

paths	description	theme
https://github.com/KarlenS/BikesploreLA/blob/ffa9ee5f0004ed94afd5190b34ab7fc7620010ec/GeoDataCollection.ipynb	- code, viz - commits made over 2 days; message about data techniques - first commit code and min output (3 iters) - markdown added, viz added - code added to existing code cells - code is reordered - clean up (merging cells?)	code comments
https://github.com/jordanosborn/CrowdAnalysisDDM/blob/81afb236a731fbf1c1c6b5ab58d9e743490c54c2/code/DataExploration/SingleScale.ipynb	- final nb is all code, one viz - all commits in one day, comments are vaguely descriptive ("added fit data", "fitting is tragic") - consistently adds code to the end, big jump due to adding only viz - noted that plot was incorrect via code comment rather than markdown (some ppl prefer using code comments as internal notes and markdown as external notes?)	functions; impromptu comments
https://github.com/rickyviank/UWA_Channel_Modelling_ARL/blob/9bf44cf1e884f4190218b3b21e4a7602db6566d3/Channel%20Modelling.ipynb	- final nb is majority code, some viz - commits over 4 days, last commit is 4 months later, comments describe goals - output changes, combining of code cells, adding print statements and more text output due to simply running cells (large amount of text output and one viz is likely spike cause) - modifying code to add more viz, commenting out majority of code cell and adding new lines to perform convolution a different way (using existing library) - tends to combine code cells every few commits, uncommented commented code to revert back to old way of convolution - added new code cells and vizs throughout nb, again commented code and tried library again with some changes and plotted, added new code and vizs to end - final commit adds code and viz to beginning, adds multiple code and viz to end (trying different methods or visualizations for same thing (noisy vs filtered signal))	code comments

https://github.com/michaelweissman/ORF_467_HW7/blob/0795137b2897260e64bc1ca0be7ba214bf16bf25/ORF_467_HW7.ipynb	- code, text output, vis, code comments - commits are made over 16 days; commit messages are automated - source code added along with viz - more code is added and output changes - lot of iteration on output, source added to the end of the file - code is also modified towards the end	code comments
https://github.com/DrClick/ARCRacing/blob/4428a244c5a4627f4550eba066657b5a87ff0602/training/prep_car_data_shadow.ipynb	all commits made within four days; final notebook has only code and vis; the commits remove and split code into smaller code cells; commits also remove visualizations; repo contains other notebooks; Visualizations disappeared/broke; commits align with data-science goals; notebook is non-linear	code comments
https://github.com/cathrineot/studioprojects/blob/e6193fc9cdb665340dab125c4cab1e504d9f3366/code/fbi/FBI.ipynb	- all commits within 2 weeks - commit comments are not meaningful ("new", "m") - first commit had some code cells and text output (data exploration, cleaning), last commit had a lot more code, viz (heatmaps, graphs), and a markdown cell - added mostly to end of notebook, very visualization focused to explain results, some additions in middle of notebook to visualize previously generated dataframes - commented out output, reduced dataframe size to clean up notebook near end, added more code cells to middle of nb and created two new identical viz (one called test), added to end of nb with some more code and new viz	code comments

https://github.com/matttrent/mnist-uncertainty/blob/0e971cdfbebaadfbbf79286991cbbbbe00fd1473/not_ebooks/edward_supervised_classification.ipynb	- final notebook has commented out code, code, visualizations - commits made within 3 days, commits correspond loosely to high level data analysis techniques - first commit contains code and visualizations - lots of code being added on the second iteration - heavily iterating on the code (adding comments, and iterating on visualizations) - reorganizing code	code comments
https://github.com/mirgeet/thesis_project/blob/ceef23aa125895afd3b974e96c18e1fac488fb56/not_ebooks/12-dfa.ipynb	- All commits made within 3 months - Added exclusively to the end of the notebook - Not a template - Added parallel techniques/ expanded on existing techniques - Commits correspond with items being added but not necessarily oriented towards programming goals - Content of the first and final commit include code and visualizations (more of both committed across versions)	code comments
https://github.com/AnastasiaSulyagina/NLP/blob/7ab69b08d6acd185f407b1e0435deecc29f8eb14/notebook.ipynb	Two different iterations of ideas being represented in the commit behavior. The first iteration modifies and add code cells. The second iteration adds code cells and leverages existing code to examine new results. Output changes from graph to text across commits. Not template.	code comments
https://github.com/abylikhsanov/advanced-lane-lines/blob/ddbd95a1a746470e447a70dfc5d73b91b46cb8a2/detect_lines.ipynb	all commits made within two days; started with only code and one visualizations; ended with only code and a text visualization(?); Iterative changes to code and output; commits don't correspond to anything. Other scripts in the file	code comments

https://github.com/anishpdalal/twitter-sentiment-analysis/blob/c6edd88f81ffd34269d3f287bff448d561f785b6/Political_Social_Media_Post_Analysis.ipynb	- code and visualizations only - commits made within 20 days; commits correspond to analysis techniques - first commit is just code and visualizations - added code to the end of the file - iterate on code and viz	code comments
https://github.com/brookefitzgerald/ESNResearch/blob/2a5db3a084bf1bf5369ecdbb5be925d7ad85d87/ZhangClasifyPlayground.ipynb	- code and visualization only (mostly just code) - commits made within a month; commit messages correspond to data analysis actions - first commit contains code and viz - added code cells to the end of the file (along with viz) - added code to the middle and added more code to the end of the file - added libraries, consolidated code into functions - iterated on code cells - removed certain visualizations	code comments
https://github.com/randomBEEAR/robustness_measures_notebook/blob/92439d1bd0de298fb4f52685c95f53bc0781fec8/plot.ipynb	- majority of updates done within a month, one last commit ~2 years later - commit comments are fairly meaningful - added markdown explanation to end early on - expanded on calculations for one aspect done in one cell to span multiple cells - used a different method for plotting (not run) and deleted old code - moved code cells around, tended to add markdown one commit after the code was committed and would edit the code as well - last commit had an extreme amount of viz (going from no viz to a lot of viz), seems like there was also code additions focused on visualizing, such as seeing a viz and writing more code to further explore or analyze it	code comments

https://github.com/gmaher/DeepBrain/blob/964e47fdb837b8f091dec6a3d449d0e4529a20e/python/dicom.ipynb	Image analysis nb. Commits made over 2 weeks. Started from code cell with comment and import statements. Added code cells, visualizations. Only nb in repo.	code comments
https://github.com/sharmasavaranan/Machine_learning/blob/e7d78872edf8530a4c7dffffc1d8fe5bad2ba49c/Data%20science-python.ipynb	only one commit via master branch, more commits if you don't switch to master - 1 month commit span, comments correspond to dates - final is mostly code and text output, some viz, some markdown - start with only code and text output - markdown are short phrases/notes to themselves (don't always match the code tho?) - adds a lot of viz at end, then deletes a majority of the nb (i think reverting back to an old version), in last commit readds all the viz	objectives
https://github.com/mhincz/counterdiabetic_plots/blob/12bfccc9fb87ac82cc3374153312379a66c3f832/counterdiabetic_plots.ipynb	- lots of code being added to the end of the file, with viz - code is iterated upon, some commits don't include changes but evidence of being executed	code comments

https://github.com/racinmat/GTAVisionExport-postprocessing/blob/396e9ab64e23c90e272c3083aad16048935bd60a/inspecting-broken-annotations.ipynb	- 3 week commit span, descriptive comments - final nb is only code cells, text output, and viz - start with some code and viz, mostly text output - makes changes/additions to code/viz in middle of nb but also end	code comments
https://github.com/crazysal/MNIST-Logistic-Regression-MLP-CNN/blob/06a199c8ba554e8d8cb026faca5fb2c6192d8843/Project%203%20-%20Classification.ipynb	- code, viz - 10 day commit span; commits correspond to models and changes being made to the notebook - initial commit contains just code and text - code being added with code comments and output - different types of tests being added - code cells deleted when they're heavy	code comments
https://github.com/aemerick/galaxy_analysis/blob/61b4865e9ed6278a7538fc41f00a7e77ac72868c/notebooks/longnormal_model_fits.ipynb	- most commits done in 6 months, last commit is 1.5 years later to update to python 3, some descriptive and some not - final has code, viz, text output - first looks similar to final version - large additions of code and viz throughout notebook, not many changes to the end	code comments

https://github.com/SourojitDas/ML1819--task-105--team-24/blob/aa7e5665a8c6313c6bf100b5dac513e39e1781604/Checkout%20Codes/Second_Assignment/ImplementationCodes/Adult%20Census/Statistics.ipynb	- final has code, viz, tabular output - 2 day commit period, slightly descriptive - start with a few code cells, table - second commit adds a lot more code, tables, and viz to end - after that, fairly small changes (either output changes or small code line/comment changes) - several commits that just switch back and forth between a version number	code comments
https://github.com/keremayoz/RetroMovieRecommendation/blob/e16b80806d2ed03ef46571c3b32c06738e3c6794/src/auto_encoder.ipynb	- code, viz - month long commit span; commits correspond loosely to the project management timeline - just code in the initial commit - code being added to existing cells and new code being added to new cells - iterating on the same code cell - Adding cells designated for output	code comments
https://github.com/andyashall/maths-for-ml/blob/56ca54829385ce5f04b30542c4fc55573ff08a/Simple%20linear%20Regression%20using%20Stochastic%20Gradient%20Descent.ipynb	- code and visualizations only - commits made within 20 days; commits correspond to data science techniques - first commit also just code and viz - expanding on one code cell - added code to the end, visualizations added along with code	code comments

https://github.com/himanshubabal/Google-Street-View-House-Numbers/blob/bead00a11d065be74cec54f22fb3537dbdd17c90/MNIST_Train-11-9-17.ipynb	- code, code comments, and one text visualization - commits made within 3 days; commit messages aren't meaningful - first commit code, and viz (commented out code and empty cells) - commented out code and visualizations are generated - code is added as commented out code - code consolidated to functions - file is completely emptied out!!	code comments
https://github.com/contemplat0r/LANL_Earthquake_Prediction/blob/88c063902c2e8532b95558841892a86d40dec475/lanl-models-with-parallel-earthquakes-release.ipynb	- code, viz - commits made over 2 days; commit messages don't correspond to anything - first commit doesn't contain any output - code cells are added and iterated upon	code comments
https://github.com/HassankSalim/CloudCarre-ML/blob/806b8df8751e7c2bf71826ea9dfdb7c30a67e82/Cluster.ipynb	Random exploratory nb. Commits made over 2 days. Started with 1 code cell with import statements. Added code cells with text and table output. One of two nbs in repo.	sections; cell structure

<a href="https://github.com/alchemists5617/Climate-
Network/blob/1f843a0b399ed91d006897c54f46d5a229a143b8/Network%20Predictive%20Power.ipynb">https://github.com/alchemists5617/Climate- Network/blob/1f843a0b399ed91d006897c54f46d5a229a143b8/Network%20Predictive%20Power. ipynb	- subheaders, viz, code; - two week commit span - all commit headers are automated - code being systematically added on each commit - headers being added along the way'	code comments
https://github.com/afbaum/ baseball/blob/ 8b013f3a4d47 56fe27460b53 fdd92a06b7fd 75d4/Baseball .ipynb	- merged code cells, iterated on visualizations (at the end of the file) - markdown annotations are added along the way - code cells being iterated on and code cells mainly being added to the end of the file'	sections
https://github.com/karino2/ ClassSim/blob /6dcba21efdc 6867aed3cd92 b37b6a5e481 1db148/GMM similarity. ipynb	- final has code, markdown headings and sentences, tabular and text output - about 2 months commit span, briefly descriptive - start with code cells, text output, markdown listing stack overflow resources - adds content throughout the nb, adds markdown as code is added - tends to delete portions of code but also add large amounts of code - adds objective (replicating results in a paper) in second to last commit along with other additional markdown	sections

https://github.com/RayGonzalez/CS166-Modeling-/blob/e0d0ea56d1b2e4db856293e53a028a20ac34bd65/FinalProject.ipynb	- code is deleted from one section - code cell is iterated on in another section - visualization is rendered - markdown and annotations added to the top of the file, some visualization unrendered - new section added with code, headers, visualization (code divided into sub-cells) - annotations are being added along the way - instructions and assumptions for subsequent interactions are added as well - iteration happens within each subsection	sections
https://github.com/ppik/playdata/blob/62612dbaf26ed139d0ced800d199b897a0f2ba01/Kaggle-Expedia/Expedia%20Hotel%20Recommendations.ipynb	- code and viz added to the end of the file - each code cell is annotated - new sections are added and delineated using markdown cells with headers - guide towards new algorithm being added - alg implementations are added with both code and viz and headers'	sections
https://github.com/wballard/kerasvideo/blob/703bdc278ada3b1ff0c718d8ba4e19036eccf4f4/ConvolutionalNeuralNetwork.ipynb	convolutional neural network exploration. started with many code cells, a few lines of descriptive text and a viz. ran a cell to get output of a graph, added a markdown cell to explain what the code was doing, changed code a few times to make viz more readable. added more markdown to describe what was happening, added relevant code. reverted back to old notebook version, so removed code and markdown. readded the removed code and markdown, added more detail to description, removed code at the end.	sections
https://github.com/ericmjl/bayesian-stats-talk/blob/279fa1668344c738c7774a54de78cc800235e896/slides.ipynb	- notebook is primarily being preped as a medium for presentation, most of the edits are markdown oriented - each cell is being evolved agnostic to the sections of the notebook - the presentation is always sophisticated, the visualizations are well rendered and code well documented'	sections

https://github.com/stevendae/CarND-Vehicle-Detection-Tracking-P5-StevenLee/blob/a9e768120782b729662fa72e08880c53c956b2ca/Pipeline_Video.ipynb	- image detection - 11 day commit span, descriptive of coding accomplishments - markdown headings, text output, viz, code - adds some more headings throughout, code and viz at end - mostly small changes throughout nb and adding new content to end in some commits	sections
https://github.com/vucko83/classification_task/blob/2a15e8a60539dbb88b9b3dee8a88d2209b417070/Classification.ipynb	- code, text and tabular output, viz, markdown headings - 2 day commit span, not descriptive - start with code and headings, vaguely similar to final nb - deletes some code, then makes changes throughout nb to code, outputs, and headings - makes large code changes in final commits	sections
https://github.com/jiagarwa/capstone-project1-Jitendra/blob/1d71e6720f53a37a9687c3331c8e9999625a967c/Inferential%20Statistics/capstone-project1-Inferential-Statistics-Excercise.ipynb	All commits within 4 days. Capstone data analysis project. Starting point has code and markdown performing an overview of data. A lot of deleting file and reuploading updated file. Only notebook in file.	sections

https://github.com/jesbot/DAT_SF_14/blob/f53b930ca556cc1e44dc65b2c889fed74a68f485/home/work/MyWork/ homework3/ Homework3. ipynb	- code, tabular and text output, viz, markdown - 3 day commit span, progress oriented ("finished first draft") - homework assignment but not templated - start with code, text and table output, and markdown analysis of results - removes commented code, finishes assignment in one commit - only two commits that make significant changes, others are removing commented code or fixing typos	
https://github.com/oxanozaep/Instacart_Recommender/blob/ed232341300612ca00a6f63f11dd4f22434526db/notebooks/EDA. ipynb	- final has code, markdown, text and tabular output, viz - 12 day commit span, a few are descriptive - makes small changes throughout nb, adds majority to end (code and viz) - adds markdown headings as code and viz are added - added more headings, viz, and code throughout notebook, also deletes some code and viz from end related to analyzing clusters - last commit is also more of cleanup, adjusts longer code lines	sections
https://github.com/whitead/numerical_stats/blob/9cdd349971806176bf581f6a69f7b98b905d5020/unit_10/lectures/lecture_1. ipynb	- headers, code, annotations, viz, conclusions - commits are made over 3 years! commit messages correspond to lesson plans - first commit contains headers, viz, code, annotations in outline format - observing modification to source code and markdown in parallel	sections
https://github.com/bmmidei/SliceCast/blob/91d8eada4238c3b23ee0e9997b91ed223410a235/inference.ipynb	- final has code, text output, viz, markdown headers - 3 week commit span, briefly describe what has been added or changed - mostly very small changes (few lines of code, changing an image, output changes) per commit - adds markdown objective header and other headers and short descriptions near end in a documentation focused commit	sections

https://github.com/teuben/SLICA/blob/644e0d2a600f2f10aade5bfc8fe2c581c0ef81b/notebooks/orbits.ipynb	- subheaders, viz, code, annotations - one year commit span - commits loosely correspond to high level features and changes being made to the notebook - initial commit contains description, subheaders, code and visualizations (relatively complete notebook?) - added pivot of the initial model (after the initial commit), markdown and subheaders are being added with each code/vis adds - code commits and small changes being made to code overtime (a year later)'	sections
https://github.com/jbarkhan/cs_example/blob/62c14d3e5f0dc41a9c2ba205048d8c2235e8ee2e/compressive_sensing_example.ipynb	- commits made over one year span (most within a few days last year) - commit comments don't correspond to anything - header, annotation, code, code comments, viz in the final version - code and viz, with conclusion annotation (initial commit) - Markdown explanations added along the way - Small changes being made to the annotations along the way	sections
https://github.com/kjwinston52/Jupyter-Notebooks/blob/80d450e6b3c7bd7954149a9482bbf92d056ce839/Chapter_10.ipynb	- super basic chapter to start out with - commit messages are all automated - commits made over the course of 2 years! - multiple notebooks in repo - code and markdown added at each iteration - previous code being replaced - markdown being edited meticulously - iterations on examples and analysis - examples of theory being outline	sections

https://github.com/JoekingCooper/btakehometask/blob/76e45a6f6982d6ac946854659c3a961bfedfd2/notebooks/8_Forecasting_Bespoke_Models_w_Errors.ipynb	- final notebook has headers, subheaders, code, visualizations and annotations - commits made within 5 days, comments correspond to literal changes made to notebook content - first commit contains visualizations, annotation, subheaders - iterated on visualizations and updated plots - 'cleaning' notebook mentioned (removed cells, added markdown, putting things into functions) - adding and modifying markdown towards the end (minor changes to the source code)	sections
https://github.com/enalisk/blog_sandbox/blob/cfdc9a1cb5959254e4825034275664df2366ffaf/Tree-Based_Priors.ipynb	added code cells/lines and visualizations. commits correspond to programming goals. (new cells and code added in each commit). Not a template. Lots of other notebooks in the repo.	sections
https://github.com/qrrble/Forex-RNN/blob/7d255a352d7eb81d43c2988515c52cff928e6da4/FOREX_RNN_-_MultiCurrency.ipynb	- initial notebook contained annotations for hyper-parameter tuning, visualizations and code - commits made within a month-long period (4 in one day and one a month later) - commits loosely aligned with data-science goals - parallel libraries and APIs being added in second commit - hyper parameters are tuned once - explanations and subheaders are added all at once	cleaning
https://github.com/jcomish/MovieAnalysis/blob/25d7ab6b91a61465b015e41195612a78d8ad7a02/MovieAnalysis.ipynb	- lots of code being added and changed - code being split and pushed into functions - lots of code reorganizations being demoed in this notebook - headers and annotations are being added along the way - sections being designated with headers - code comments being added too'	sections

https://github.com/gperrin12/bike_lane_pothole_cv_detection/blob/eb1b1935bd33a80f1e2a7609be2642de03ebc27c/neural_networks_definition_detection_111517_GTP.ipynb	- code, headers, code comments, viz - commits made within 18 days; commit messages aren't totally meaningful of 'evolution' - first commit has fewer viz but all the same highlevel elements - code which hasn't been finalized get pushed to the end of the notebook under a non-descript title; annotations are notes the users are taking - code added to the end of the file - headers specify models which are being tried (headers, code, and visualizations are being added at once)	sections
https://github.com/GuannanShen/image_data/blob/ecc782931c47751846b037804d516d2918521cc2/bio_image/GaussianFilter_CannyEdgeDetection.ipynb	- final commit contains visualizations, code, annotations, headers/subheaders, code comments - commits made over the course of two months (loosely), commit messages loosely have to do with data analysis - first commit contains markdown, visualizations, headers/subheaders - expanding existing code cell and adding more cells to the end (with visualizations) - parallel analysis added with headers and annotations (all iterations) - basic iteration also checked in'	sections
https://github.com/nkasmanoff/MLstuff/blob/503eba35429d1aaa1fe668a5588c5272a7683a4c/Phys476/project/SpreadGuesser_4-23.ipynb	- code, code comments, headers, tables/text viz - commits made within 16 days; commit comments correspond to data science workflow - first commit contains code, annotations, code comments, text output - code cells are being modified, markdown being added - libraries added - code cells being moved around - minimal annotations added - duplicate cells added and modified	sections

https://github.com/mayankkaiizen/jupyter-notebooks/blob/8c48cfd72eb63bc22463047106dc3df765486817/ML%20and%20DS/Pandas%20-%20Working%20with%20Time.ipynb	- final has code, mostly markdown headings, and text output/dataframes - commit span of 8 months, 3 about a week apart, last two several months apart - commit comments are not coding goals ("regular backup") - initial has code, image, and markdown headings and short explanations of code and output - adds a lot of code, text output, and some markdown headings throughout nb - very few if any changes in last two commits	sections
https://github.com/annekehdyt/text-transparency/blob/72c0376c21e26431f8d4b846ff2b3907ea70e2a5/Dataset%20load/Amazon%20Review%20Data%20Load.ipynb	- code and output added in the middle of one labeled section - all changes made within one commit - minor code changes towards the end, to the middle of the notebook in different sections'	sections
https://github.com/jjpeh/W207/blob/4385a567818f172ace500878551e76c280c27e9a/W207_Home_Price_Prediction.ipynb	- viz, headers, subheaders, annotations, code, code comments - commits span a month; do not correspond to anything - data import code from kaggle (headers/subheaders are added along the way) - headers/subheaders added before code - annotations are added along the way	sections

https://github.com/Ngort/CS156/blob/b7d2a4f485923b28a9c4b5cd2f32f2dce67913da/LBA/ML_LBA.ipynb	- final has intro and analysis text and headings, code, vizs - all commits in one day, comments are vaguely descriptive - first commit has some headings and code cells - second commit nb is extremely similar to final commit nb (added a lot of code, headings/text, and viz to end of nb) - basically cleaned up nb (added headings, removed bad or unnecessary code)/made small code adjustments/added a viz in rest of the commits	Objectives
https://github.com/belindazeng/goingviral/blob/9d3db7168aa45c569447f87be2a888019293be07/goingviral.ipynb	- twitter content analysis school project - final has code, markdown, text and tabular output, viz - a lot of commits by multiple authors over a week, descriptive of changes made, some are merge conflicts - start with a lot of content (markdown, code, tables, viz), completed markdown intro, initial data gathering and cleaning, has markdown headings and empty code cells to signify what needs to be done - bulk of additions are done to end of nb, but there are small changes made throughout nb as well - some commits are small changes (adding a comment, changing a few code lines) - large jump to explanatory may be due to addition of multiple viz and a little markdown (filling in some of the todos) - in later commits, there's more focus on going through the entire notebook and cleaning it up/making changes	Objectives
https://github.com/zeronoun/general_assembly_dsi_project/blob/c6b5f872d6e8d76ba171ed2a2221e9ce5f3dc246/Project%204.ipynb	- 2 day commit span, not really descriptive of coding goals - web scraping project - code, text and tabular output, markdown - majority of work is done in first commit - second commit follows up with most of the remaining work (adding more regressions) to end mostly - other commits make small changes throughout the nb	Objectives

https://github.com/Nikronic/Coursera-Machine-Learning/blob/253864691d65388b4e9891e35c5d3e3fb9182ffe/Week%206%20-%20Regularized%20Linear%20Regression%20and%20Bias%20v.s.%20Variance/Regularized%20Linear%20Regression.ipynb	- annotations and code are being added in parallel to the end of the notebook - sections are added to the bottom of the notebook - markdown is edited through out the commit window'	Objectives
---	--	------------

https://github.com/agalea91/nhl_2016_playoffs_twitter/blob/df1e71a5e91ab8de5a2d2d1185ea925629d6afa9/nhl_playoffs_2016_twitter_study.ipynb	- commits done within a month - commit comments more descriptive of progress than programming goals - nb starts with code and markdown headings and descriptions, a clear plan of action - reorganizing, adding, and deleting code cells - added to "analysis outlook" with observations and how they conducted analysis - clearly designated "test" code using markdown or comments - added more markdown and ran cells to produce vizs, added more viz - more code reorganization, modified graphs to include more info, a lot more viz, markdown, code, and output - added more code, all output removed (but score seems to have increased to be more explanatory??) - used code comments to discuss doubts or concerns ("don't trust this plot") - final commit added some lines of code (to serve as checks? length, printing history), added more viz, commented out several blocks of code but did not add or iterate on it	Objectives
https://github.com/amogh112/detectingFacesInWild/blob/90015da37e63b91dae2ba7dbf946cecc0da55b5e/data-analysis-Copy1.ipynb	- headers, code, visualizations, minor annotations - 4 month long commits; commits messages correspond to data science techniques - first commit contains a list of TODOs, minor headers/annotations, and code (minimal output) - code added to the end of the file, cells modified to generate better viz, headers also added along the way - added parallel technique (?)	Objectives

https://github.com/eclarson/DataMiningNotes/blob/d36923731a764ee246e4e343a3cba780ff3c8ab4/04.%20Logits%20and%20SVM.ipynb	- markdown, code, viz, images, text output - 1.5 year commit span, descriptive - looks like a logistic regression walkthrough by a professor, first commit seems complete so further commits are simply updating it (often to match with things like python 3) - changes/adds code and some markdown, heavily comments, changes in output throughout nb	Objectives
https://github.com/moonman239/Capstone-project/blob/530555f06765cd1d8cb54d9ceffd21de6ec9a218/Capstone%20project%20prep.ipynb	- code and markdown - 1 week commit span, descriptive - each commit contains a relatively small amount of changes (a few lines of code changed or added) - iterative process of changing some things, running and getting errors and committing and then fixing those errors in next commit - because author makes a few specific changes per commit and commits are labeled well, it is easy to understand what is changing - attempted a different method of doing something by deleting old code and adding new code	Objectives
https://github.com/MrGeislinger/UdacityMLND_Capstone/blob/22348ccf8f55950249f1a53d4fec98b24d00acf6/asl_recognition.ipynb	- most commits in a week, last commit almost 3 months later, comments say which model was run with what accuracy - image recognition - code, markdown, viz - start with pretty solid outline, has code and headings, later cells not run - makes changes throughout nb, deletes some code and produces output, third commit only adds some text output and apparently leads to huge explanatory spike - mostly running code with different parameters but also changes code/markdown throughout nb at times	Objectives

https://github.com/ck-unifr/yelp_dataset_challenge/blob/17617bc8544f118ea5d8947ca61610398aa67fdd/notebooks/predict_star_ratings_with_photos.ipynb	- most commits in 3 days, a few more about a month later, not descriptive - final has code, headings, tabular and text output, viz, images - starts with 1 sentence objective, markdown, code, table - makes changes throughout nb but majority is added to end - does not execute cells in one update, may be reason for spike - adds some more detail to markdown in last commits	Objectives
https://github.com/katyhuff/npre412/blob/c84f84ad8546f7cf3541121d6bd7c98748aa2d80/enrichment/enrichment.ipynb	- markdown and code changed in conjunction - markdown heavily added'	Objectives
https://github.com/isenilov/WetRoad/blob/18392c84332660f115a40c9a05d8872f4210517f/ae.ipynb	Started with markdown instructions. Code and output being added. Not a template. Multiple notebooks in repo. Commits do not document changes in the way the data is processed-- even though that does happen between commits.	Objectives
https://github.com/peinrules/Darin/blob/a3a34261141bdfa5bd212dcacf5fac41679bf7b/labs/lab-4.ipynb	started with only markdown; added code (cells and lines) for each markdown section; commits based on markdown docu sections; added headers and more code at the bottom of the doc (course template)	Objectives

https://github.com/shuojiaشي/women_in_stem/blob/21e85d9d7424ae2d0162c32c5ea294c7b0f0ce49/notebooks/Women%20in%20STEM_project%20report.ipynb	- women in stem analysis project - final is incredibly well organized, has table of contents and headings, code, text and tabular output, viz - most commits in 6 days, last commit 8 months later - first commit is entirely markdown (headers, table of contents, intro text) - added a lot of code, markdown, and viz in fourth commit, basically almost completed final nb - commits before that made very small text changes, and commits after it make relatively small changes as well	Objectives
https://github.com/aalto1/digitalepidemiology/blob/58d20dc156dc0c1de6689604cbe2888f6e226099/buratti_2.ipynb	commits made within 19 days; messages don't correspond to anything; code and markdown added in parallel in entire new section in the middle of the notebook; iterated on existing code too; sections at the end of the notebook being expanded on as well	Objectives
https://github.com/erayon/India-Australia-Cricket-Analysis/blob/df054ddc207e56fca9a4a97b8fe57dcade03da40/bowling_comaprism%20.ipynb	- final has code, tabular and text output, markdown headers, viz - commit span of 5 days, not descriptive - starts with a heading that states overall objective - makes code and viz changes throughout nb, markdown headers are adjusted and added as code and viz are added	Objectives
https://github.com/lionfish0/dp/blob/b7c430cb25321f3cdc14b1aba2724bb203065035/1d%20Kung%20selecting%20hyperparameters.ipynb	All commits made within 3 days. Notebook exploring how to select hyperparameters. Starting point was some code blocks with text output. Added more code blocks and also markdown cells with insights. Many other notebooks in file.	Objectives

https://github.com/lemoinef/practice/blob/96d3401eca96b0d034ef9387044aa9af0b93faec/dlnd/Week%203/sentiment_classifier.ipynb	all commits made on the same day; commits correspond to project goals; started with high level goal descriptions; headers added and code segments for new sections of analysis; consistent commits corresponding to different goals; linear development of the notebook	Objectives
https://github.com/redlix/BigDataPractice/blob/90087401971d701b573958457eeb753ca762a1b3/KaggleCompetition/kaggleCompetition.ipynb	template, libraries are pre-loaded, (subsections, annotations, code) added within one commits. No system of committing. Visualizations added all at once. The rest of the submissions are output iteration or bug fixes.	Objectives
https://github.com/jennakosborn/mae622/blob/33daad39cf102d08c75edc25ace36415c16c3e1c/Lesson10.ipynb	- commits occur over a week - commit comments focus on progress ("almost done") - first commit is just a header and code cell of imports - very steady addition of markdown descriptions/thought process and code - changes are consistently done at the end of the notebook - final nb is mostly code and markdown, some viz	Objectives

https://github.com/Cronjaeger/almost-infinite-sites-recursions/blob/046f57afda921e13d559c981f0aa0740468fc73/ipython_notebooks/ComputationaI%20Experiment%20--%20Recombination%20vs%20Recurrence.ipynb	- annotations, headers, subheaders, viz, code - full year of commits - notebook starts with a description of a plan and code - markdown being added and code added to the end of the file - some code is commented out - annotations are broadened - code is added along with annotations - adding different types of visual analysis at the end of the file - 'experiments' are added to the end of the file - linear progression in analysis being seen	Objectives
https://github.com/nrnatesh/bioe131/blob/4953cc95e91daea22f2899188cdd3c66a55c9eaa/lab6/bioe131_lab6.ipynb	biology genome sequencing analysis. first commit was majority textual output, had several markdown cells and code cells as well. cleaned up output, added many code cells and produced numerical outputs for each, added viz and a markdown analysis of the results and more code cells. while no explicitly stated objective, this is for a lab and it looks like the author had something they were specifically looking to accomplish. not a template bc no todos or explicit instructions, at least visible on the nb.	Objectives
https://github.com/jmeisele/kaggle/blob/ca8ef3ba13fcef32d45747d0931f3956cdcfd01d/Marketin gFinalRealEstate.ipynb	- output, objectives and descriptions added to the notebook - subsections have an outline descriptor too - code comments direct objective of each cell - code cells and visualization added - code comments are used to designate functions'	Objectives

https://github.com/makennedy626/Thinkful_Unit_4/blob/4ba4d026add2950a665a0b1e4b710ba55f4aef68/4.2.5%20Challenge%20Boston%20Marathon.ipynb	- headers, annotations, code, visualizations, code comments - commits made within 10 day; commit messages correspond to dates - first commit contains annotations, code and viz - more code added - code added with code comments (to the end of the file) - replacing bulky code with APIs/ simpler techniques - render better visualizations over iterations - linearly lay out multiple iterations and visualize each	Objectives
https://github.com/dsalgado_r/mathematics-for-bigdata/blob/4b1ba34487f0d504b14a00a752455a039fb16e73/Deliveries/Delivery3/Deep-Neural-Network-for-Classification.ipynb	all commits made within two days; started with code/visualizations and a high level explanation of the goals around the notebook analysis; added subheaders, code and visualizations; (big changes) -> (little changes, annotations) << repeat (2); Some commits are based on big milestones and some are just 'saving'. Data files and R files in repo.	Objectives

https://github.com/pussinbot/tAlko/blob/9ea226b3c43a55e657da0d6c125efcbcb50ef196/notebooks/learning.ipynb	- AI q-learning - final has code, images, markdown, text output, viz - commit span of 2.5 weeks, comments are more personal notes ("check if over done", "wow training works") - start with markdown defining elements that will be used in the code, code, some viz, markdown headings throughout initial nb so author seems to have a plan - consistently adds to end of nb, at one point adds a code cell with commented TODOs ("whats next" commit) - tends to use code comments instead of markdown for personal comments ("everything works but slow", "might want to add..") - not a lot of markdown and all is in top half of nb	Objectives
https://github.com/freelanceastro/galactic-machine-learning/blob/f4b0e22341da21341128f1d2e6771daf8432dd6c/Galaxy-Zoo-Machine-Learning.ipynb	- new sections are added with markdown and code - iterate on viz - sometimes more code added than markdown - iterating on existing code cells, adding code to the end of the file too (with markdown) - markdown clean up phase is added as well'	Objectives
https://github.com/ssimontacchi/1906-Traffic-Patterns-Market-St-SF/blob/523e416f7bc830369c341d4c68b60cf0cb0efa48/mapping_market.ipynb	- header, subheaders, code, code comments, visualizations - commits made within a week; messages correspond to data analysis techniques - first commit seems like a half filled out template, but the code cells are relatively empty, the goals are set and are pretty clear - reorganize structure of the notebook - headers highlight the goal of each code section - unusable code pushed to the end of the file - random and empty code cells added between each code	Objectives

https://github.com/JeffMacaluso/Misc/blob/9caeb767c4c3c12af15887157fe8d5b7b9137f11/MachineLearningMOOC/Ex7%20Python%20%20k-means%20Clustering%20and%20Principal%20Component%20Analysis.ipynb	- each section is initiall expanded using markdown cells - code is added with some markdown or modified markdown / code comments - each section expanded with code, heavily annotated with code comments - markdown and more source code being added'	Objectives
https://github.com/adjakka/Mini_Projects_Data_Science/blob/2bb75855b1bafa05d62a4648eac91390606619d3/Unsupervised_ML_Customers_Segmentation/Mini_Project_Clustering.ipynb	- in the beginning, certain code cells are expanded and visualizations are rendered - some markdown stating directions are removed - code cells are added to the end of the file - subsequent changes are minor and made to markdown/viz'	Template
https://github.com/zya/notebooks/blob/719380e87b822e3223405f67c1bb541c7c8c5721/udacity-deep-learning-foundations/language-translation/dl_and_language_translation.ipynb	Training neural network for language translation; changes based on big project goals; starting point was a list of code cells, markdown comments and code comments asking for function definitions to be modified; modified only code within existing code cells, did not add or remove cells or markdown, added output; LOTS OF notebooks in repo (course template)	Template

https://github.com/hj940709/Deep-Learning/blob/648846aaacac291d97112ecf21ab930181152f97/image_project.ipynb	- code, annotations, headers and subheaders, visualizations - assignment, template - Commits made within 8 days; commit messages correspond to dates - first commit has annotations, visualizations, headers/sub, code (unexecuted) - code cells are added and iterated upon - markdown cells added with code - visualizations are iterated on - discussion added to the end of the notebook - code comments are added - iterated on markdown	Template
https://github.com/Wolfgang90/dlnd_language_translation/blob/9b3ea6e74bf33c3bb1e6d952e1361d8f0412a04/dlnd_language_translation.ipynb	- same language translation assignment as above (but follow slightly different paths!) - final is again majority text output and some code and markdown - majority of commits in 1 day, comments say what has been implemented - follows a very sequential path of adding code to fill in TODOs, just longer bc author focuses on little code in each commit (commits are broken down by functions implemented) - end exploratory dip is probably due to large addition of text output	Template
https://github.com/nathanquinteiro/ADA-Homework/blob/f10a72ea787432d00112eb4728bd836ca6380f2d/01%20-%20Pandas%20and%20Data%20Wrangling/Data%20Wrangling%20with%20Pandas.ipynb	- template - commits made within the same day - commits correspond to homework exercises - added very few code cells and markdown - systematically adding code and markdown - reordering cells (all at once) to match assignment requirements'	Template

https://github.com/byam/dlnd/blob/3fc897c0dfc2c553c0aee1f5a6dc7a40d200ea61/project2-image-classification/dlnd_image_classification.ipynb	Project for image classification neural network. Starting point was many unexecuted code blocks and markdown. Executed codeblocks, edited markdown cells to answer questions, edited codeblocks. Only notebook in file. Repo with many other project files.	Template
https://github.com/tcrack326/DLND-Language-Translation/blob/c1759cc3b69ee71871294383edbeab28bc5c94a8/dlnd_language_translation.ipynb	- language translation assignment: templated, has TODOs - 1 week commit span, not very descriptive ("updates") - final is majority text output and then code and markdown - tends to add code and run cells sequentially to fill out TODOs, sometimes comes back to earlier cells to change some code	Template
https://github.com/CheloGE/MachineLearningProjects/blob/961d147b1bfa207c40bb978b5767b84473b23619/customer_segments/customer_segments.ipynb	- headers, subheaders, annotations, code comments, visualizations - assignment (TEMPLATE) - commits made within a week; commits correspond to questions in the notebook - TODOs needing to be filled but the first commit is a full template - some and markdown being added (minimal) but for the most part, the source code in code cells being modified - extensive markdown description added regarding results and procedure	Template

https://github.com/connermc/image-classification/blob/2c715f0791acff54a297b42b250e637629c9166f/image-classification_dln_image_classification.ipynb	- markdown modified in the beginning - code added to existing code cells (rarely adding new code cells) - hyper parameter tuning involves small changes within commits (done towards the end)'	Template
https://github.com/hey-simone/blog/blob/f32fea74588cb676c4278e474c50953d49ce53d7/projects/Forecasting%20Stock%20Prices%20and%20Generating%20Buy%20Sell%20Signals%20with%20the%20Python%20Data%20Science%20Ecosystem.ipynb	- subheaders, headers, viz, code, code comments, annotations - commits made over 3 months; commit messages are automated - code comments and markdown are modified - code deleted and markdown added (cleaning?)	Template

https://github.com/OlgaBelitskaya/data-analyst-nd002/blob/e7d539731a92495a886a848b4f1e6d77358115d3/Data_Analyst_ND_Project5/P5_Identify_Fraud_from_Enron_Email.ipynb	- headers, subheaders, code, code comments, annotations, viz - template - code comments added - commits made within 7 days; commit messages are nearly all automated - parts of the notebook deleted - lots of iteration on output, source code cell iterated upon and markdown text added to existing markdown cells in parallel	Template
https://github.com/luappay/Projects/blob/7e8b2bfdc57b7d744e46c5a3c82139eae0aea732/Project3/project-03-starter.ipynb	changes are made across one month; changes correspond to high level data-science goals; code changes --> insight annotations; template; markdown and code added for answers	Template
https://github.com/shotaro12oyama/udacity-vehicle-detection/blob/a5774fe1513addb631ea2c354eb2031a3e06b5a8/Vehicle_Detection.ipynb	Commits made over 2 days, most in 1 day. Vehicle detection project. Starting point was cloned template. Changed code blocks and added/edited some markdown. Only notebook in file.	Template
https://github.com/MartiRubio/VA/blob/3e453c8079a135179034f15a9fcec91e049e26fa/VA_P5/p5_2018-2019_2solve.ipynb	all commits made within 10 day span; full notebook committed; code cells filled up; code cells added; visualizations added; commits aligned with goals within the notebook	Template

https://github.com/smhilde/ut_hits/blob/ca23f9fe34f97ebc0524e6356c20c6eb16b8f270/ACC%20Aim%201%20Description%20of%20Hit%20Exposure.ipynb	Analysis of sports data for a football team. Starting point was a project with markdown questions, and some codeblocks with table output and messy visualizations. Added code to generate more visualizations. Edited markdown to answer questions. Only one of 2 notebooks in file.	Template
https://github.com/carlespols/DSCI6001-student/blob/25b2406d74fe2f4ec0386f65497bcbbb6a017427/week1/1.4/DSCI6001-1.4.Lab-CarlesPolesMiengo.ipynb	worksheet using numpy to work with matrices. more markdown than code at beginning (markdown directions), large addition of code cells and one markdown, few comments added, more code cells added at end	Template
https://github.com/jorgergb/sentiment-network/blob/ff52253c538039eb2dd01f4e87c4ed8ef211b4c8/Sentiment_Classification_Projects.ipynb	sentiment classification for neural network interactive tutorial. lots of code and markdown text in first commit. filled in TODO parts of code, lots of output/changes in output for all other commits	Template

https://github.com/rpicatoste/dlndf-project2-image-classification/blob/14a53457d31b8f39155ff7306f27c7a3bc447f1d/dlnd_image_classification.ipynb	Lots of other script files in repo. Training image analysis neural network; almost identical labels to nb 336. Class oriented submissions -- three different goal captured. Commit/Iterate cycle.	Template
https://github.com/xrick/first-neural-network/blob/37fe008d55d5ac52a43cb849160d30c4a45f15e4/Your_first_neural_network.ipynb	commits made in a 20 day span; commits don't seem to correspond to anything within the notebook; one/two giant commits and others only iterate/modify existing structures	Template
https://github.com/anandha2017/udacity/blob/2d84ea6f22ff37135e89a7b5d84cfd92f644641d/nd101%20Deep%20Learning%20Nanodegree%20Foundation/DockerImages/08_sentiment_analysis/notebooks/Sentiment_Classification_Projects.ipynb	all commits made within four days; template with code/markdown; adding code to each block over the course of each commit; added code cells (minor); commits do not correspond to project goals; the iterations are in-built in the notebook; final notebook has more visualizations	Template

https://github.com/anarchistMegaByte/SDCND/blob/e68651a0a8636946729b0f59cd0faffc52acd5d4/CarND-LaneLines-P1/P1.ipynb	- templated image identification - commit span of 6 days, vaguely descriptive - start with given template (code, markdown) - fills in all todos in early commit, spends rest of commits making changes to code and viz (improvements, fixing bugs) throughout the nb	Template
https://github.com/saiabhishekgv/dog-classifier/blob/ccc6b1fc7d718928f837804efdbb66fe9da37bc5/dog_app.ipynb	Template with comments being filled. Code being committed based on programming and data goals. Some code cells added towards the end. Updates markdown cells with better answers. Added visualizations and removed text which was direction oriented (rather than insight oriented).	Template
https://github.com/WeiShiwei/cs231n-Spring-2017-Assignments/blob/3a9291f5489f80e3d8b2b58cd753699652e1ae08/assignment2/TensorFlow.ipynb	model training using TensorFlow interactive tutorial. started with a lot of explanatory and heading markdown and code cells. added lines of code throughout last commits, deleted some lines as well	Template
https://github.com/akivanov/SK_NLA2015/blob/35b25420b08c2e637fa002c88c82194df0abc3d8/ps2/pset2.ipynb	- annotations, headers, code, visualizations, subheaders, assignment oriented - commits made within two weeks, commit messages correspond to assignment - template - markdown full filled out, code cells are empty, except for one - expand on one code cell and visualize from that cell - fill out next cell, with similar visualizations - markdown is modified with code' - iterate on visualizations	Template

https://github.com/dbbrandt/project_plagiariism_detecton_dbb/blob/2873f2b90e7340ef26ff054f58497a8dd9cbf39c/2_Plagiariism_Feature_Engineering.ipynb	- feature engineering project - final has markdown, code, images, text/table output, and viz - 1 week commit period, comments report completion progress of the project - first is mostly markdown and code cells that need to be filled in - runs given cells which produce a lot of table or text output, not that much code added - delete commented code, add more code and a few new outputs - just sequentially goes through nb from start to end, likely starts high and ends low bc starts with a lot of markdown and more code is added than markdown by this author - addition of two graphs at the end didn't seem to really impact the score	Template
https://github.com/abramhindle/keras-deep-learning-tutorial/blob/ba8a998a04ea7c6b74209657457cced980f855d4/Keras%20Deep%20Learning%20Tutorial.ipynb	- final notebook has headers, annotations, subheaders, code, and visualizations - commits made across a year and a half; commits have loosely to do with data science tasks - first commit pretty much has all the same elements as the last commit - subsections have clear objectives (objective oriented notebook) - code cell which are added are clearly templates with implemented solutions - solutions are being added to the end of the file - markdown and code added	Template
https://github.com/markdyousef/machine-translation-project/blob/d137060cff5d5c67bcb41254e8531e096ca6d6ea/machine_translation.ipynb	- header, subheaders, code, text output - assignment (TEMPLATE) - commits made across 4 months; messages correspond to data analysis workflow - initial commit has no output, and TODO statements within each code cell - code and visualizations keep getting added within the template construct - new code cells do get added, doesn't seem like markdown gets added too	Template

https://github.com/lizhang101/CarND/blob/7b6d2c1a7ad0126d4c6026df9f7a54155bfa7fa/CarND-Traffic-Sign-Classifer-Project/Traffic_Sign_Classifier.ipynb	- added code cells within each section, iterated on code in existing code - merging and splitting cells is real'	Template
https://github.com/sshhhan/reflections/blob/86a516086da2fd2484edec7b641507ef5ed71038/cn-deep-learning-master/tutorials/sentiment-network/Sentiment_Classification_Projects.ipynb	all commits made within a day or two of each other except first commit (~1 month away); interactive tutorial, first commit is just pretemplated markdown and todo code cells, continuous filling in todo code and producing output.	Template
https://github.com/abhikbanerjee/Machine_Learning_And_Deep_Learning/blob/7e0d77dfa8848b2747798f8dc9b58ddf06a1f47f/Python_Machine_Learning/src/06_Model_Evaluation_Hyperparameter_Tuning.ipynb	- header, subheader, table/viz, code comments (minor annotations) - commits over 10 days; message correspond to data science workflow (loosely) - First commit only contains one empty cell - adding code and some viz - code being added to the end of the file with code comments stating the purpose of each cell - subheader added as the notebook is expanded - overall header added mid way - new visualizations added to each subsection - validation added to the end	cleaning

https://github.com/footh/nuclai/blob/3650dd712828d2139daa7d51775b84c65f140874/notebooks/image_aug_sandbox.ipynb	- image manipulation - final has code, text output, markdown, images - 1 month, comments briefly say what part of code/process is being worked on - start with code, text output, images - adds markdown once to explain what is happening/needs to happen in a code cell - fairly consistently adds code (and runs to get output and images) to the end of nb, changes a little code in earlier cells - miniscule changes in last several commits (changing location of images, changing a variable, commenting out some code)	cleaning
https://github.com/MacYeh/Udacity-Machine-Translation-Attention-Pytorch/blob/f6ea2a11467167e3bfa6151dd3f98cd3314993262/machine_translation_pytorch.ipynb	changes made within a week; commits correspond to big goals being met; all the commits are showing iteration; more code is being added so the score went down; not a template	cleaning
https://github.com/chrishokamp/dl4mt_exercises/blob/2ce7f3606cc19fb9a8ed18dff7dcf2995c17cc3f/notebooks/day1/theano_logistic_regression.ipynb	- commits done within 3 days - commit comments are basic but descriptive - seems like a logistic regression tutorial but has TODOs, appears that TODOs are things the author wants to keep track of rather than provided by the tutorial bc TODOs change or get added as commits progress - todos and some code gets removed as they get accomplished - some commits have a small amount of changes, a little code added, a few lines or output changed, addition of todos - replaced prior code with new iteration/idea - adds markdown and a viz near the end - definitely greater focus on cleaning up the notebook for others to use near the end, the beginning and middle were exploration based	cleaning

https://github.com/WoutAarts/Advanced_House_Pricing/blob/1c808b74c0e6fdc8f12885919fbc85450ec1731a/advanced_housing.ipynb	- subheaders, code, viz - two week commit span - commit correspond to literal additions to data science analysis (histogram, markdowns etc) - adding code along the way with sparse markdowns (mostly headers)'	sections
https://github.com/johnociocoiu/udacity-and-self-driving-car-project-5/blob/b15146fb3169b142dcc9387ab8def8ea8d5866d3/vehicle-detection.ipynb	Image analysis tutorial. Commits made over 10 days. Started from code cells with visualizations. Added code cells, markdown headers and descriptions, output. Only nb in repo.	cleaning
https://github.com/cnehrkorn/CaliVoting/blob/80bd1cb1656cd2a5ef38d69a0b14fd8106852ea5/explore_CA_state_wide_measurements.ipynb	Data analysis tutorial. Commits made over half a month. Started with code cells no visualizations. Added code cells, markdown with directions, high-level visualizations. One of many nbs in repo.	cleaning
https://github.com/isgiddy/oc4920_code/blob/a4a7e67a01e99b9c31b6fccf48fdbcf73ea7354b0/Week1_Exercise2.ipynb	Data analysis exercise. All commits made over 2 days. Started from code cells and table output for data overview. Added code cells, headers, and high-level visualizations. One of 3 nbs in repo.	cleaning

https://github.com/kirnh/CarND-Vehicle-Detection-P5/blob/844360df3a5222c06a3fdb0c7159e0ebd40298ea/main_notebook.ipynb	- 1 week commit span - commit comments are somewhat descriptive - first nb is all code, some image output - deleted code cells and code lines (potentially things that were tried but didn't work out) - changes throughout the nb (adding code cells/lines), not only the end (but consistently adding code to end) - code is well commented - near the end, "clean notebook" commit adds headings, numbered steps, deletes empty cells and test code - reorganized code cells (training a model), but most of the code was not changed - very small changes made after "clean notebook" commit - the author tended to use empty code cells to break up sections before adding markdown in the cleaning phase	cleaning
https://github.com/pedroallemnerez/WebTraffic/blob/c2ce093c9221e2a46d2a2d6659540138684ea721/traffic_analysis.ipynb	- final notebook contains visualizations, code, headers, subheaders and annotations (even conclusions) - all commits made within 15 days, high level descriptions of what is being committed (no real commit message) - first commit contains visualizations and code only - libraries being added, new techniques added to the end of the file - moved code around - adding headers/subheaders to sections which have solidified - added annotations (observations) - added detailed text descriptions - adding more complex visualization code , iteration on the visualizations - adding 'conclusions' section	cleaning

https://github.com/kdubovikov/alcohol-research/blob/d3eff87bb0656a0d0ffc51d95c086b40854fa739/student-analysis.ipynb	Data analysis nb. Commits made over 4 days. Started from code cells with text, tabular output and visualizations. Added code cells, markdown headers and descriptions, high-level visualizations. Only nb in repo.	cleaning
https://github.com/stkbailey/nashnetx/blob/454754afcd e8afa5287b41b61101f647de d6d751/4_Nashville-Relationships.ipynb	Tutorial on graphs. Commits made over several months. Started from short code cells with error and tabular output. Added code cells, headers, text description, and visualizations. One of many nbs in repo.	cleaning
https://github.com/mayankagggarwal/MyLearning/blob/56e78f57b922e395efc446d8609b44fffb81bd7e/articles/24Tutorials/3_BigMartSales/3_BigMart_Sales_Prediction.ipynb	Data analysis nb. Commits made over a month. Started from one code cell with comment. Added code cells with output. Only nb in file, but many files with nbs in repo.	cleaning
https://github.com/NicoleLam/Data-mining-experiments-NYC-data-on-Kaggle/blob/9e3f6051c2d7066f36c5163b00fb761a16ea588b/Untitled2.ipynb	- subheaders, viz, code - a year long commit span; commits correspond loosely to program management goals - title of the notebook is not set - initial commit, only code and viz - markdown headers allow for annotating after all the code/viz have been added - changed code pipping - more markdown added as code is added - a lot of the code gets deleted and annotation gets added towards the end	cleaning

https://github.com/IrisQY/CME249-Assignments/blob/bb0eaf90c030cebbcf8d0aa48ee0c5c8d68942c2/MiniProject.ipynb	Short data science project. All commits made in 9 days. Started from empty nb. Added code cells and headers. One of many nbs in repo.	cleaning
https://github.com/alexandermlcocq/python_student_debt/blob/2280e0347a80a692cb4b0cbe467da20364be980f/Project01.ipynb	- subheaders (some annotation), code, viz - one week commit span - commits correspond to literal additions to data science analysis components (markdown, headings, bug fixes, data..) - additions of code and markdown added in an intersperced way'	sections
https://github.com/efinkel/Claustrum_analysis_notebook.ipynb	- most commits done in a week, final commit one year later - commits somewhat describe progress/goals - deleting code and adding new code, output, and viz - often single code lines for printing or verification - experimenting with visualizations - added markdown headings about halfway through, removed not working code, added to viz code and to code at the end for creating dataframes - added new code cell to try an alternative approach, didn't run old code cell - modified vis, reorganized code cell order and renamed markdown headers accordingly, removed unnecessary code - cleaned up prior experimental code, added new code (beginning of nb), final commit only has output for cells that had not been previously run	cleaning

https://github.com/acesaif/Juice_Recommender_System/blob/03fdbeb397afb73829e29b5d337a97669c3fe66/ml_juice_recommender.ipynb	- subheaders, viz, code - two week commit span - commits correspond to actions taken in the notebook - code and tables to start out with - adding visualizations and code at each commit, and headers after the sections have been added'	sections
https://github.com/nmningmei/intuitive_probability_games/blob/98aa46d3dedeb92bbcf2346b3492b9b9972f3f99/10%20side%20die%20problem.ipynb	- final nb has code, markdown, viz - commits over 2 days, comments meaningless - first nb is a few code cells, next commit runs cells to get viz and adds code to get another viz - adds heading and markdown analysis sentence, code cells, and viz to end of nb in each of last two commits	cleaning
https://github.com/ctslater/gaia_experiments/blob/74ba1aed5c9dac9e5600d2afb59b50de7f99ce08/model_mu_b.ipynb	- all commits made within two days - starts with only code and viz - subgoal being achieved in one day, improvements to the model over the commits - new sections of analysis added with subheaders (linear)'	cleaning

https://github.com/aleks091/machine-learning-capstone/blob/2f1b621c17a8f44032e5fbf6729a3b6ca001ac4e/InceptionV3.ipynb	- all commits done within 4 days - commit comments are not very descriptive - start with exploratory nb: mostly code cells, lots of print statements and empty cells - next commit cleans up, removes print statements, adds more markdown and adds to existing code, removes a lot of prior code at end of nb - small code changes, more markdown addition in final commits, adding some code to end - final commit (also called second version) has more code, markdown, and images added throughout nb but also added at end of nb	sections
https://github.com/alexing/spotify_rewrapped/blob/03502b1303aa4722d8cdc9354476bf3662e89a5d/wrapped.ipynb	commits made over 15 days; markdown and code (cells) are being added in conjunction; commit messages aren't really meaningful; code is being added pretty linearly	sections; cleaning
https://github.com/kimrharpner/image_segmentation/blob/bad4488e6321f1f55a44124d24ca719a2eb38bcf/edge_detector.ipynb	Image filter project. Started from code cells with image and text output. Added code cells, markdown headers and brief intro, visualizations. One of several nbs in repo.	cleaning
https://github.com/doolingdavid/IBA/blob/620c5d200d6f967506cbcca43930365e813632af/MAKEALLPAYMENTS.ipynb	- final nb is code, descriptions, and viz - commit span 2 weeks, not descriptive - initial nb has two code cells, just imports - next commit has a LOT more code, markdown explanation of the data, some viz - from there, fairly small changes, a few code cells added to end of nb, execution count change, adding a few more code lines and a markdown line	cleaning

https://github.com/yuhsunc-hien/ML_final/blob/27284cab1a9e1ab7d3040f2fc7d753a280d85373/ML_EM_PCA_ndims.ipynb	- code, viz - week long commit span; cannot understand commit messages - initial commit has code, viz, commented out code - adding annotations in markdown cells, entire parallel modeling technique in one cell	sections
https://github.com/mycarta/fun-with-king-tut-catscan/blob/2159d7b8bf87263a48001df764f00d08a71de287/decimate_and_interpolate.ipynb	- all commits in one month - commit comments are somewhat meaningful, some just say update or finalized and others discuss accomplishments - final commit has significantly more code, viz, and markdown (first commit has no markdown) - second commit only added markdown with descriptions and result analysis - steady addition of markdown explanation and code cells - uses markdown to write down future goals at the end and can understand the author's thought process ("maybe it would be easier to..." "how do i compare..")	sections
https://github.com/hamogu/Cepheids/blob/e2f26510631682b6cd94b602e0aa01b84c70fb00/snaps_hotprogram.ipynb	- lots of commits, made over the course of 4 years - commits correspond to experimentation goals (moving code, creating autonomous code utilities) - started out with only code (not data science) - Adding lots of code comments along the way (addressing self) - Adding markdown insights, and new techniques - Adding visualizations (obvious exploration) - Alternate methods being added to the bottom of the notebook - Notebook used as a scartch pad, scaling elsewhere - subheaders added - isolating code to function (cells -> functions) (cleaning up)	sections;cleaning

https://github.com/AnnemarieMcDonnelltemp/data-set/blob/32bac2e011937d6ad38da1990cc17bcd70a82/Data%20set.ipynb	- final has a lot of text at the beginning, majority code and outputs and some viz - 2 week commit span, comments are not very descriptive but note errors ("code to work fail", "trying to create table") - first nb has a few code cells but majority error output - next commit only has 2 code cells and a df output (more explanatory - is it really?) - fairly steady addition of code to end, changes throughout, spike is due to many visualizations produced and a long markdown intro added to beginning - exploratory pitch as author adds code to try and create a table in sql - iterative process of trying things out and commits with errors then a new commit with errors fixed - headings and more text added in second to last commit	cleaning
https://github.com/mshabram/PyStan_Mass_Radius_Eccentricity_Kepler_Exoplanets/blob/6a6e48332adbb0b72ab73f59093eb5a7d052a34d/codes/MRE_PyStan.ipynb	- commits occurred over almost 2 years, three "cycles" - commit comments do align with goals/provide valuable info - last commit has a lot more code, output, and viz than first - replaced and filled in model info (changing from one model to another), added multiple lines of code to end of first cell, ran last cell - changed input data, committed fixes and working output - second cycle, more data exploration in middle of nb, printing dataframes and lots of output, more code and output at end of nb - another code cell added to middle to define a dict, a lot of new visualizations formatted properly - markdown explanation and many code cells and output added to end - commented out lines and removed code cells throughout nb (changes in output as well) and added new code cells to end	cleaning

https://github.com/Azhag/Bayesian-visual-working-memory/blob/6410f0ae50aae71661d3efde8dbde64d87213bcf/Experiments/dualrecall_fitmixturemodel/model_fitting_interactive_try.ipynb	Commits made over several months. Data fitting notebook. Starting point was a few codeblocks with errors or output. Added more code blocks with visualizations and some markdown with insights. Many python scripts in file.	sections
https://github.com/jaidmin/machine_learning/blob/6c6fdb65ce10a040d8e1071a38ab41bf1b8869e6/machine-learning-ex3/ex3/machine-learning-ex3.ipynb	- shifting cells - adding new source code to the end of the file - changing and iterating on visualizations - general iteration - no other real changes past the first commit'	removing visualizations (adding more code)
https://github.com/numb3r3/3/Applied-Data-Analysis/blob/0701471e88bb6bb7ac735388f37b480f50e1c964/talking_data/notebooks/TalkingData.ipynb	data analysis - cleaning, training, testing. first version is simple train and test, has code and markdown cells (questions, analysis of results, headings, graphs). editing markdown cells, adding code and removing code cells. all output and markdown cells removed (added some comments in code). added more code, reorganized notebook into fewer cells of related code.	removing visualization and markdown (adding more code)