Repository and Mining of Temporal Data
Milestone 5 Evaluation

Team Members:

Jessica Nguy	jnguy2014@my.fit.edu
Siomara Nieves	snieves2014@my.fit.edu

Faculty Sponsor and Client:

Philip Chan	pkc@fit.edu

Meetings With the Client:

Feb 26, Mar 19

Progress of Current Milestone:

Task	Completion %	Jessica	Siomara	To-Do
Prioritize Q3	90%	60%	30%	Fix plotting on Django
Narrow Data	75%	50%	25%	Finish analysis with different granularity
Showcase Documents	100%	50%	50%	N/A
Target Variable Search	30%	20%	10%	Django view, display of results
Upload more files	20%	20%	0%	Find more files
Q2	85%	20%	65%	Fix Fig. 3 (Cross correlation)
Improve Current Code	80%	40%	40%	Finish analysis with different granularity, ensure compatibility with Django and Python
Poster	100%	50%	50%	N/A
Evaluation Document, Presentation	100%	50%	50%	N/A

Discussion of Each Accomplished Task:

Q3: Q3 was prioritized during this milestone. Q3 is finished in the programming side. It takes the correlation values of the top-k datasets and uses those numbers to find the linear regression to find the weights. The weights are then plugged into the equation $w_0 + w_1x_1 + w_2x_1 \ldots =$ forecast, where $w_{1\ldots x}$ are the coefficients, w_0 is the y-intercept, and forecast is the value at the next timestamp. The values x_n are the values of the correlation from Q2. The value calculated from the linear regression formula is then graphed to the next timestamp on a graph of the target variable. We are using 80% of the data as training material and 20% of the data for testing purposes and comparing our algorithm results with the actual data to ensure our algorithm is correct.

The visualization for Q3 is one graph and looks very similar to the first graph of Q1. The code graphs the values of the target data with respects to its timestamps, and then calculates what the next timestamp is using the datetime API. The code will then display what the next timestamp is in an alternate color to draw emphasis that the value is being predicted and not actually part of the dataset.

Further testing is needed for Q3, as there needs to be a wider range of data to determine how close the predicted value is to test it. Furthermore, NarrowData must be adjusted in accords to the changes in the test. We are debating on how to display how close the prediction is to the actual value, whether or not it will be displayed as an R-squared value on the graph or on an additional window.

Q3's calculations have been added to the Django framework, but there has been a problem with trying to display the graph to the user since there has been a compatibility issue between the graph's JSON object and the type of the values to plot. This issue will be fixed before Senior Design Showcase and Milestone 6. As described before, Q3 will also display the accuracy of results to the user to see how close the team's prediction was to the actual value either next to the graph and in the same page or in a different window for the user's further analysis.

Narrow Data: The tag search function has been deferred to Milestone 6 and/or dropped entirely. The current milestone is focused on speeding up the analysis time and being able to handle different types of dates if the target variable has a different granularity. Currently NarrowData only accepts data whose dates are in the daily format (Y-M-d), however some data is in the monthly format (Y-M), and some are even in the yearly format (Y). NarrowData must be adjusted to be able to analyze yearly data with yearly, monthly data with monthly, and daily data with daily. Currently there is no plan to analyze data across types, such as analyzing daily data with monthly because of the loss of data when doing monthly-vs-daily.

For the website, there's the idea of asking the user whether it is daily, monthly, or yearly data by showing a drop-down button while uploading the file and selecting the case; the default database linked to Django will add a new column in order to save the selection and for a quicker NarrowData analysis.

Showcase Documents: The document due during this Milestone is the Showcase Poster and has been completed. It has been submitted for review by the faculty sponsor, and will be submitted to the Harris Student Design Center once the sponsor provides more feedback.

Target Variable Search: Has been deferred to Milestone 6. The goal of the target variable search is to provide a search box to Data Consumers that do not wish to upload a file for analysis. The search would accept a search variable and would return a list of the data, its names, description, and tags. The plan is to have target variable search done for the Student Showcase.

A search bar has been implemented on the homepage of the website and is currently being developed in order for the user to be able to look through the target variables that are saved in the SQLite3 database within Django; furthermore, the plan is to allow the user to also download the public .csv files with data.

Upload more files: Currently searching for a variety files to populate Database and ensure that Upload.py can be able to upload the data and that NarrowData.py can analyze the different granularity of the data. In the repository currently is stock data, search results history and some birth data analysis. Each of the categories is in a different granularity; stock data being daily, search results history being monthly, and birth data analysis being yearly. We plan on adding natural disaster data such as hurricane occurrences/intensity, earthquake observations and sports game data.

Q2: Q2.py is finished. The class calls Q3.py, passing through to Q3.py the correlation matrix calculated from NarrowData. The matrix contains the values of correlation, the name of the variable, the file name of the variable, and the lag number it corresponds with. There was an issue passing through the function using a numpy multidimensional array, so using pandas' DataFrame structure was used instead.

Q2 is almost done in Django, it currently shows 2 out of the 3 graphs the team has planned to show the user; the missing graph is Figure 3, Cross Correlation bar graph of lag times. In this case, there has also been a problem with the compatibility between the graph's JSON object and the type of the values to plot. This issue will be fixed before Senior Design Showcase and Milestone 6.

Improve Current Code: Code is under constant improvement. Upload.py has added to it different parsings for the types of granularity; this section relies on the user putting the right tags to the data rather than the code automatically detecting whether or not the data is already in a Daily/Monthly/Yearly format.

Poster: The poster is finished and awaiting review by the sponsor, then it will be submitted to the Harris Student Design Center and put onto the project website.

Evaluation Document, Presentation: The presentation and evaluation document is finished and written with Google Slides and Google Docs respectively. Once finished the files will be put onto the project website.

Discussion of Contribution for Each Task:

Jessica: Planned, coded, and finished writing Q3.py, including linear regression analysis and visualization. Edited Q2.py to call Q3.py and restructured results of Q2.py so that its correlation calculations can be passed through to Q3.py. Searched for more files to add into repository. Added different granularity upload options on Upload.py. Planned out a framework for NarrowData.py to be able to analyze different granularities aside from Daily data and a framework to implement target variable search. Contributed to writing the Showcase Documents, the Showcase Poster, Milestone 5 evaluation, and Milestone 5 presentation.

Siomara: Wrote Q3's app for its visualization to be shown on Django, added NarrowData to the back-end of the website using the framework's SQLite3 database, implemented a search bar on the navigation bar of the current pages, wrote part of the queries for the target search variable on Django, created users' view when displaying the search results, added Q2's code to the framework, implemented 2 out of the 3 visualization for Q2's results on website using JSON objects and the mpld3 library, created new templates, rewrote some code on Q2 for it to be compatible with Django and JSON. Contributed to writing the Showcase Documents, the Showcase Poster, Milestone 5 evaluation, and Milestone 5 presentation.

Plan for the Next Milestone:

Task	**Jessica**	**Siomara**
Website	Create more pages, front-end design	Create more pages, front-end design
Target Variable Search	Variable search on dummy database, make sure it works with actual database	Search bar, search view, user download of files
Save Results in account (provider/user)	Add database table for saving user credentials	Create Django app, add database table for saving user

		credentials
Optimization	Improve current code	Improve current code
Test/Demo System	Test cases and analysis	Test cases and analysis
User Manual/Demo Video	Write User Manual, record Demo Video and submit to Harris Student Design Center	Write User Manual, record Demo Video and submit to Harris Student Design Center
Evaluation Document, Presentation	Write evaluation document, Create presentation, put code on GitHub repository.	Write evaluation document, Create presentation, put code on GitHub repository.

Discussion of Each Planned Task:

Website: The current website consists of the homepage, the upload page, the upload/analyze page and the results page. After Q2 and Q3 are finalized, the plan is to create more views for the user such as About, Instructions, and Contact pages for the front-end. The overall design for the homepage is done but graphics will be added to make the site more user-friendly. The views for searching variables and user account will also be added once the back-end is finished.

Target Variable Search: Implement the search of target variables within the system. Currently, the website consists of an inactive search bar since testing and implementation is being developed. Allow users to see answers to Q1, Q2, and Q3 without having to upload a file.

Save Results in account: Create users' settings for the creation of accounts and logging in. Allow users to link their uploads and results to their account, create the account holder view and dashboard for quick navigation of the system. Create the private setting (public or private) when account holders upload their .csv files.

Optimization: Fix remaining issues with Q2 and Q3's graphs, aim for accuracy, optimization, and speed. Double check for accuracy results with more files and more test cases. After some changes, the system is working at a faster pace than before when analyzing a file with over 500 entries of data; check if changes in the code can make the visualizations faster to load. Make website more user-friendly so users are able to quickly and easily navigate through the system.

Test/Demo System: Navigate through the system, upload files, upload and analyze files to see if any errors happen. Check if every feature created works the way it is supposed to and record the results for further inspection.

User Manual/Demo Video: Write user manual following the course guidelines. Show how to use the system’s features, description of the written code, and example of the different types of users (provider vs consumer); also, record the demo video showing the main features of the system. Submit the two to Harris Student Design Center after receiving feedback from sponsor.

Evaluation Document, Presentation: create evaluation document stating what was completed during the Milestone and final requirements for the project. Create Presentation and upload documents to the GitHub repository and update links on page.

Sponsor Feedback:

Task 1:

Task 2:

Task 3:

Task 4:

Task 5:

Task 6:

Task 7:

Task 8:

Task 9:

Sponsor Signature: ____ Date: ___________

Sponsor Evaluation

- Sponsor: detach and return this page to Dr. Chan (HC 322)
- Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

Jessica	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Siomara	0	1	2	3	4	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10

Sponsor Signature: ____ Date: ___________