

Repository and Mining of Temporal Data

Jessica Nguy
Siomara Nieves

Dr. Philip Chan

Progress of Current Milestone

Task	Completion %	Jessica	Siomara	To-Do
Q2	90%	90%	0%	Django Integration
NarrowData	70%	70%	0%	Refine method of NarrowData, finish parsing and ensure integration works
Database Processing	80%	50%	30%	Populate database with uploaded files on website
Meta-Data, Meta-Data input	70%	50%	20%	Link to Q2 on website
Improve Existing Code	95%	80%	15%	Time Handling
Website Q1	85%	0%	85%	Redirect to Q1 app
Documents	100%	50%	50%	None

Discussion of Each Accomplished Task

- Q2
- Narrow Data
- Database Processing
- Meta-Data, Meta-Data input
- Improve existing code
- Website Q1
- Evaluation Document, Presentation

Q2

“Why was there a significant change and what are the top-k variables that have affected the target?”

- Displays a graph of the target variable and the top-k values
 - Displays a graph of the correlation value of the target variable and the top-k values
 - Displays a graph of the correlation value of the top-1 variable and its lags
- Current figures use data from the stock market, so everything is highly correlated with one another.

Figure 1

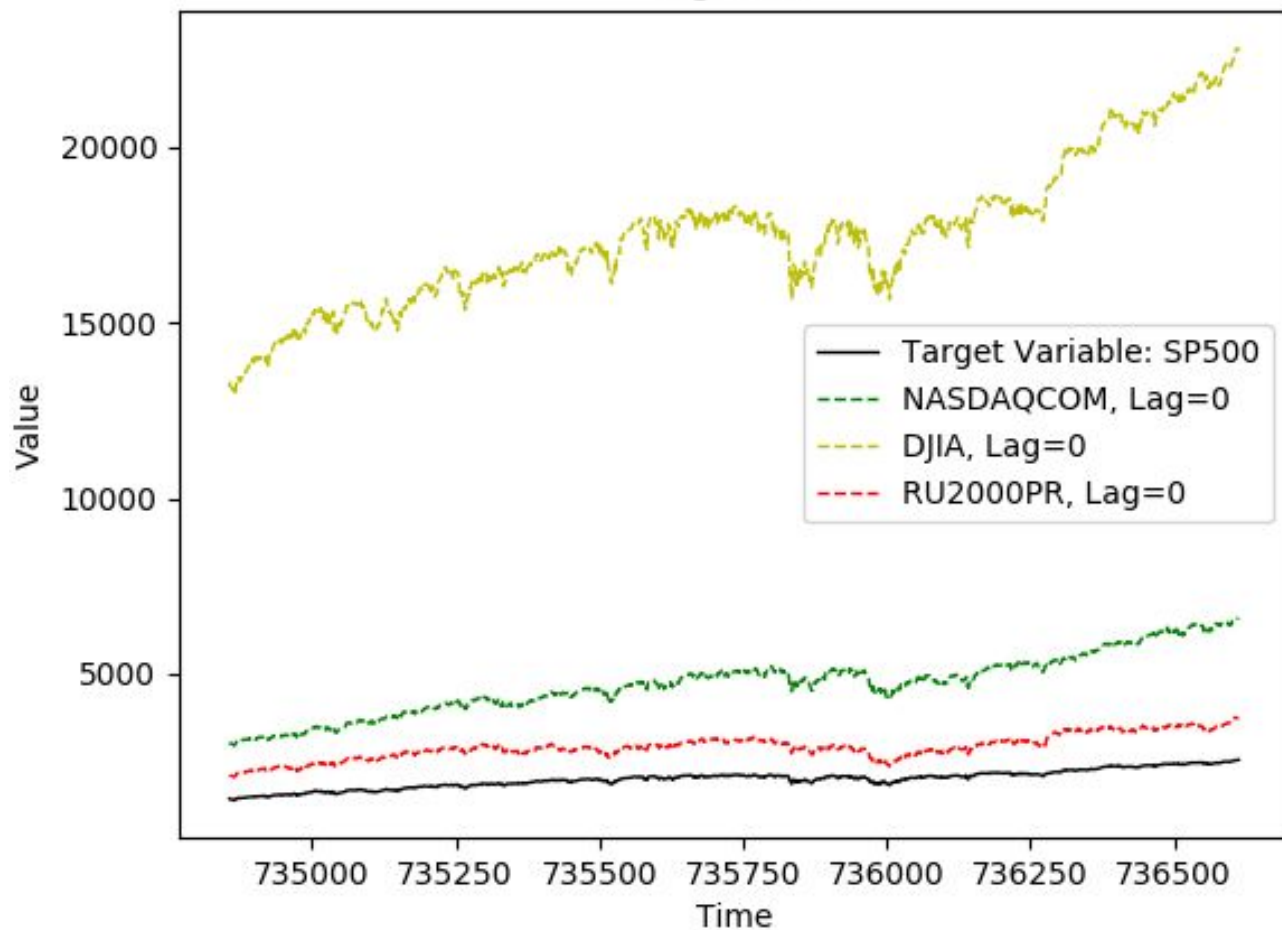


Fig.2: Correlation value of Top-K variables

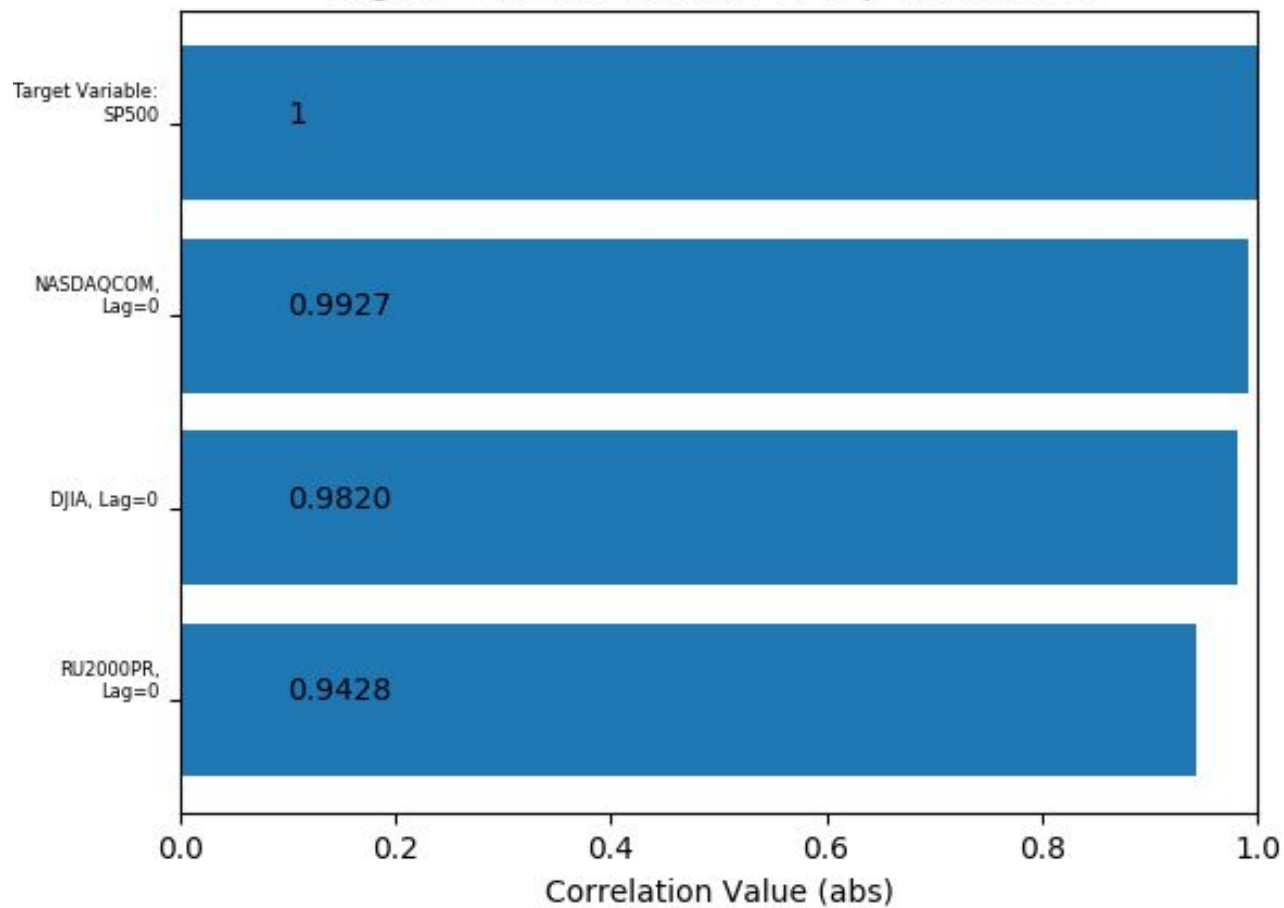
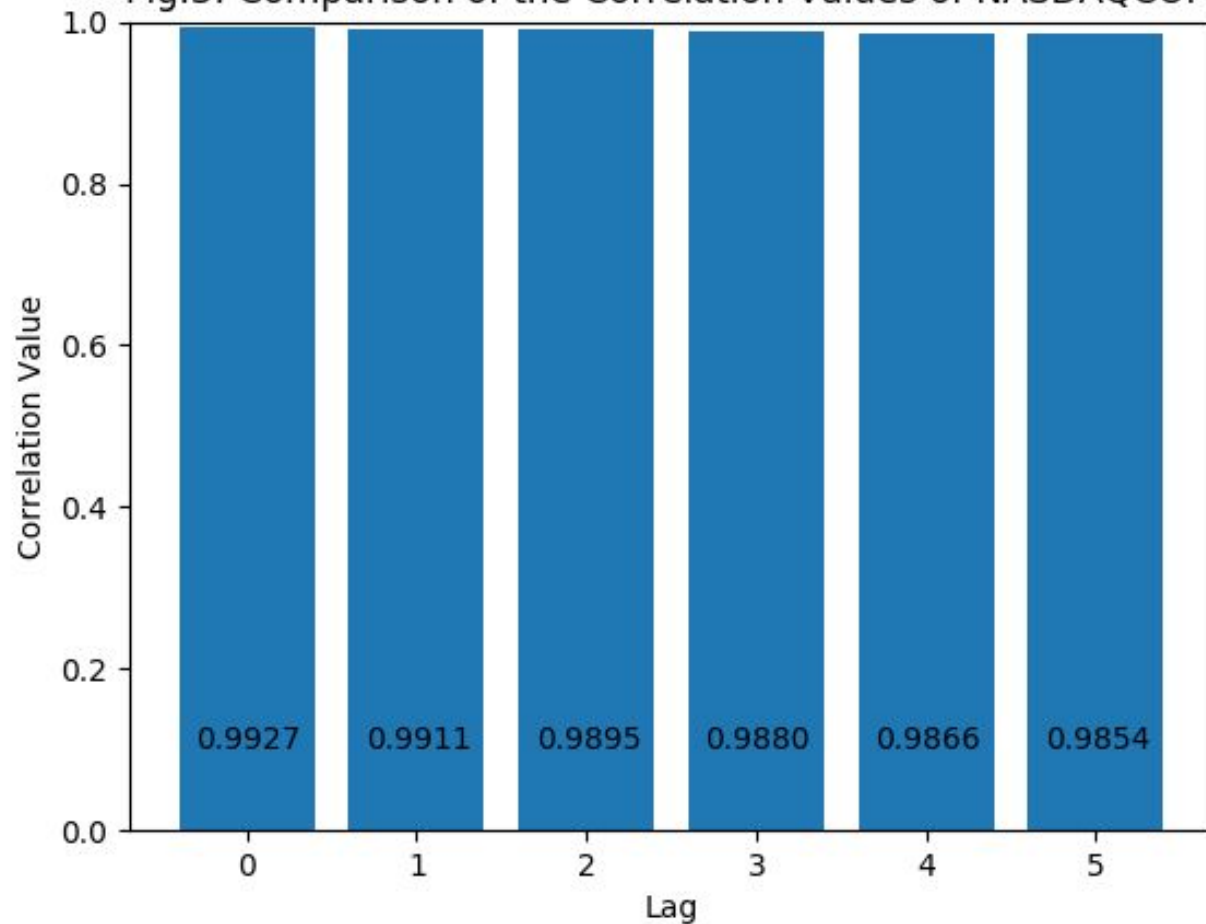


Fig.3: Comparison of the Correlation Values of NASDAQCOM



NarrowData

- Accepts data that has the same start time, end time, and number of items.
 - Rolls data forward
 - Loses some data; will be fixed in Milestone 4
 - Input.py is where users input target variable and search tags
 - Upload.py is where users upload their csv file and add meta-data to the database.
-
- Currently does not access tags to analyze target variable.

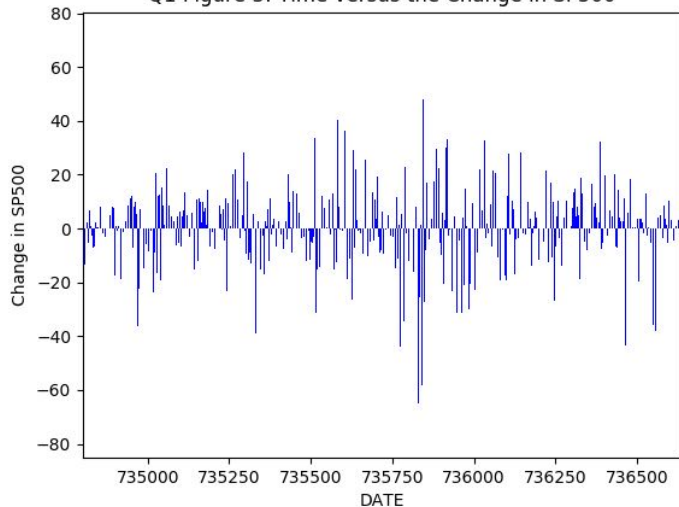
Database Processing and Meta-Data

- Uses the SQLite3 module for Python
- Schema for Database:
 - description (name, info, IDnum)
 - granularity (start, end, type, itemNum, IDnum)
 - privacy (user, setting, IDnum)
 - tags (tag1, tag2, tag3, tag4, IDnum)
- Integration between Input.py, Upload.py, Q1.py, and Q2.py does not exist at this point in time

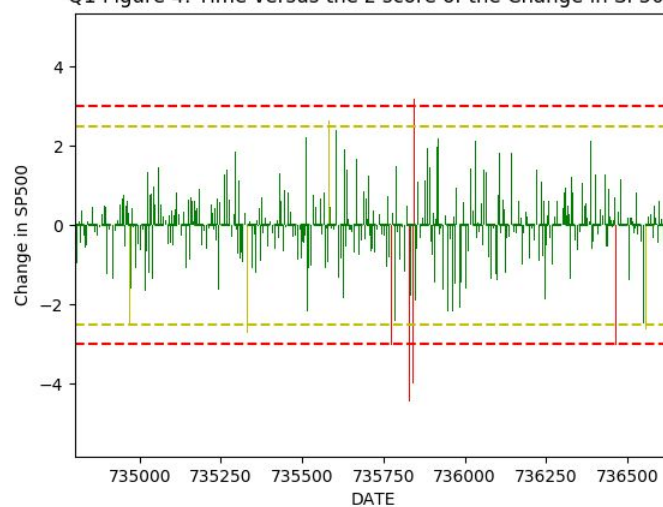
Improve Existing Code

- Bar graphs to display change between points rather than the trend
- Code cleanup
- Catch cases for Q1

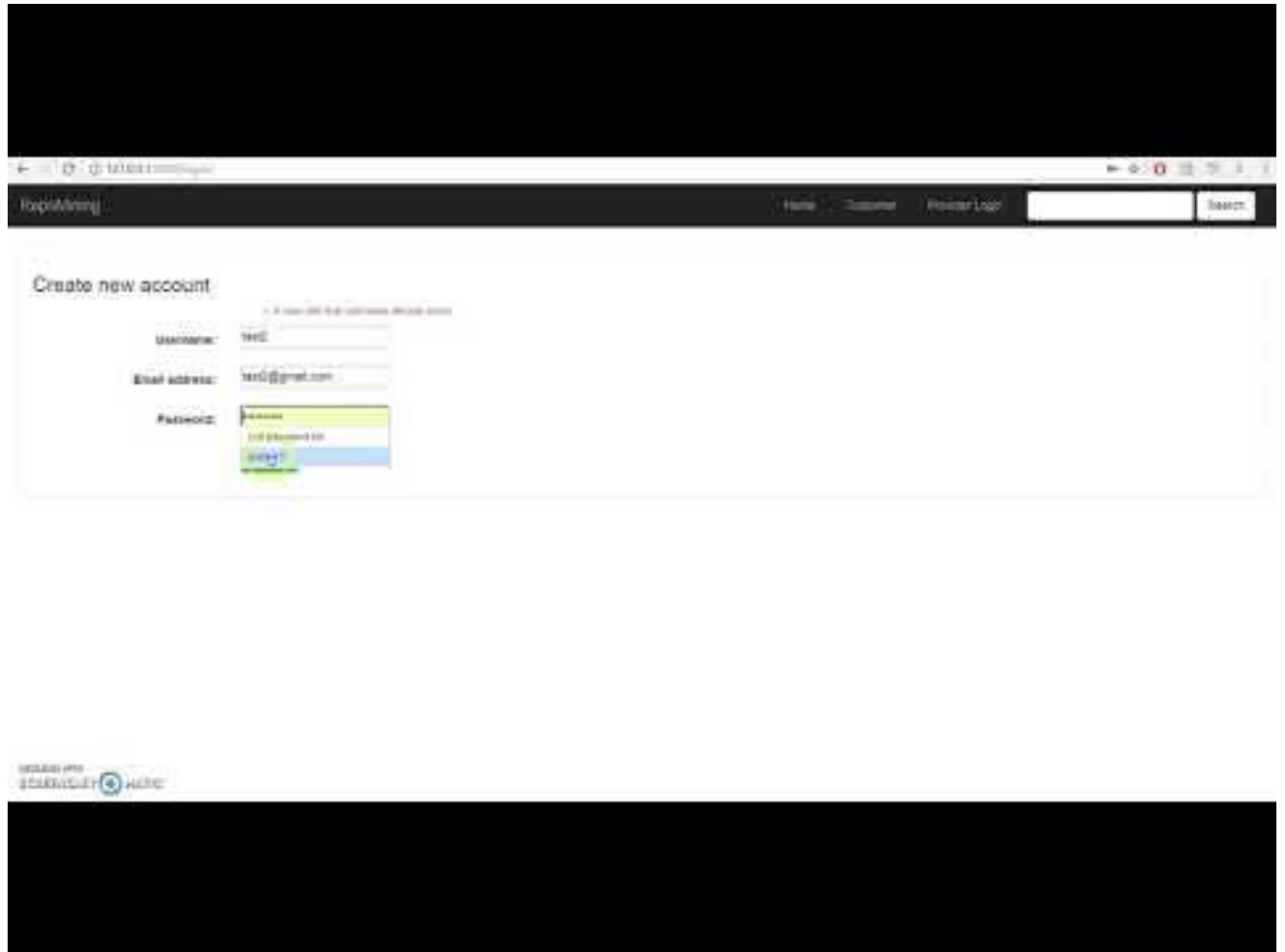
Q1 Figure 3: Time versus the Change in SP500



Q1 Figure 4: Time versus the z-score of the Change in SP500



Website Demo



Plan for Next Milestone

Task	Jessica	Siomara
Q3	Brainstorm, Pseudocode, Research if needed, Program Q3	Q3 app on Django, research, testing
NarrowData	Finish NarrowData; Figure out how to handle multiple checks for csv files	Link to Django
Django Website	Research methods of hosting code; not on LocalHost (AWS/PythonAnywhere/etc)	Create app for Q3 on current website and redirect after answer Q1 and Q2. Provider Login/Logout, Customer view
Save File Uploads	Find more csv files to populate database	Save files locally on a directory for testing, more research on saving files elsewhere
Database Input handling	Reine Database processing to be faster	Django database handling for meta-data inputs from user for Q2
Improve current code	Z-scores for Q2 Fig 1	Test cases for website
Project Plan	Write Plan	Write Plan
Evaluation Document, Presentation	Write evaluation document , Create presentation, put code onto GitHub repository.	Write evaluation document , Create presentation, put code onto GitHub repository.