**Interim Evaluation (15 points) [Soft Deadline: 1st March 2022]**

“Team Money”

Gary Ng

Wong Chee Weng

**Bottom-line:**

* We have a clear idea of our project goals and the raw data in hand
* We have used the pro-mern-stack-2 template to upload 2/4ths of our raw data into mongodb and projected the datatable on the localhost:3000 website (yet to upload to github, but in progress).
* Next we have to upload the remaining data, make the website interactive, and make the source code our own.

1) Class Participation (2 points): Participation in the class discussions on MS Teams will be evaluated. Cases where the student writes an explanatory post for the peers will be considered.

* We have posted an introduction to our problem on Teams
* We commented on other’s projects where we feel we can add-value.

2) Problem Statement (5 points): Novelty of the problem, challenges involved in solving the problem, Is the problem relevant in 2/5/10 years?, complexity of the solution, and understanding of the problem domain will be evaluated.

*Problem Statement:*

* Data is messy. Quantitative researchers working with financial data require a defined framework to ingest, organize, and process multiple datasets so that they may present research results in an accurate, informative, and interactive manner.

*Novelty of the Problem:*

* We believe the problem of data “over-supply and under-processing” is an acute and growing trend across many industries. Not only in financial data where one of our teammates has sufficient domain knowledge to drive this project, we observe similar trends in the following examples (and many other areas):
  + Multiple Personal Communication Platforms: Instagram, Whatsapp, Telegram, Signal, Messenger, iMessage, WeChat, Discord, Teams for Personal, etc...
  + Master’s in Computing at NUS: Teams for School, Discord, Luminus, EduRec, LeetCode, Kaggle, Steven Halim’s personal webpage (IT5002), Coursemology (IT5001), Kattis (IT5002), etc…
  + New Platforms: Straits Times, Channel News Asia, Today Online, Asia One, C.N.N, B.B.C, New York Times, Google News, Yahoo News, Sky News, etc…
* In the examples above, we illustrate a similar trend in personal communication, schooling, and keeping up with current events (news). In all three cases, and to our dissatisfaction, there remains multiple platforms that cumulatively provide an overabundance of data be it personal messages, school files, or simply the latest news. The bottom-line is that there is time and energy loss simply trying to aggregate data and information for simple daily tasks. This trend is similarly exhibited in the supply of financial data where there is an overabundance of data providers, but few people providing data interpretation and analysis services.

*Challenges Involved in Solving the Problem:*

* For Gary, his challenge will be to contribute to the software development portion of this project as he is new to any programming language that is not Python. He will be able to contribute on a leadership front driving the broad ideas for this project, but his challenge will be add-value in a timely manner on the programming front.
* For Chee Weng, his challenge will be able to understand the problem well as this is not his professional domain and to anticipate which portions of the project take priority over others (what to do next). His add-value will be to drive the technical development of this project.
* From a collective perspective, the challenge for this project will be to communicate well and early with each other such that both parties have sufficient time and confidence to jointly deliver our intended solution.

*Is the Problem Relevant in 2/5/10 Years?:*

* Yes, in fact we believe that this problem will grow worse if left unsolved in the coming years. Repeatedly approaching data research projects without a structured framework, data ingestion process, and processing pipeline will lead to (1) a slower overall development process (2) sub-optimal application of one’s company personnel (talent), and finally (3) the possibility of erroneous research findings and results.

*Complexity of the Solution:*

* The solution to our problem is not at all complex. On the contrary, our goal is to follow the “simple is best” approach such that an entry-level data analyst may understand the basic workflow and framework easily. Rather than complexity, we aim for a simple, scalable, and broad enough solution to not only solve this problem but provide a framework for all data projects in the future. Ideally, our solution would be to design a software architecture that provides a skeleton for a new quantitative researcher who has just joined our hypothetical company to quickly get up and running and develop in.

3) Solution Architecture (5 points): Wireframes/Figma illustrations/Information architecture/Mock Up can be considered towards this. Modularizing the overall solution and architecting the solution.

* Back-End Solution: Dataset Ingestion
  + Data Source 1: Market Data (From Bloomberg)
  + Data Source 2: Our Hypothetical Portfolio Data (From Eclipse Risk Management System)
  + Data Source 3: Research Idea from Analyst #1 (Intropic.io Data Provider)
  + Data Source 4: Research Idea from Analyst #2 (SmartKarma Data Provider)
* Middle: Quantitative Researcher’s Workspace
  + We would like a part of the pipeline to be an environment where the research may query data, apply basic data analysis techniques, and output his research results
* Front-End: Informative & Interactive
  + The final product will be a dashboard that draws upon the raw data and research results to display the relevant data (both useful raw and processed data) in an insightful and interactive manner. We will describe what we mean by interactive in the coming sections.
* Note: The above data sources are just for illustration and we are more interested in the data “pipeline,” i.e. to demonstrate how four data sources can be used together to present coherent and useful information.
* To add as we go along: Wireframes / Figma Illustrations / Information Architecture / Mock-Up

4) Legal/Other Aspects (1 point): Use of open source code in the project as well as the possibility of open sourcing the project itself need to be explored. Ideas for protecting the project from being copied will be evaluated.

* This data may be shared with our classmates and in private settings if we anonymize and change some of the underlying data. For example, we may change buy recommendations from Analyst #1 and #2 to sell, and our existing portfolio of 100 lots in S&P500 Index (for example) to 50 lots.
* We should not open source the source code of the project, but rather display to the public how we believe data should be processed and visualized under our “framework.”
* This approach will add hiring-value for Chee Weng as he explores his potential career path (note his upcoming internship with a financial services provider specializing in Crypto) and will be an interesting trading tool for Gary as he looks to become a more hands-on quantitative trader soon.

5) Competition Analysis (2 points): Explore the nearest market product that competes with your solution.

* There is a list of about 100 financial data providers, yet what we are proposing is to develop this framework as skillset for our personal toolkit. There are a few financial data aggregation and interpretation services out there, but again, if one were to sign up for even 10 of these providers, there would be an overwhelming amount of data. For these service providers, there is also the legal issue of (re)-distributing their data, which again obfuscates the underlying raw data (often companies can only show you the result and not share the raw data with you). Hence this project is important for us to take ownership of the entire front-to-end research process and improve our hire-ability in the job market.

6) You are encouraged to add a list of features, Project Schedule, and details of the project modules assigned to each team member.

The Dashboard

* We would like the home page to be line items of trade recommendations from Analyst #1 & #2 with OUR added value analysis.
* Feature 1: The columns should be able to be sorted and filtered

Graphical user interface, application, table

Description automatically generated

Feature 2: Charting

* After sorting and filtering for line items of interest, we would like to click on a line and have a pop-up chart with very simple data (daily price history and daily volume traded).

Chart

Description automatically generated

Feature 3: Output

* Finally, we would like to be able to shortlist and export the line items that we find interesting. This will either be in an “add-to-basket” button or checkbox, following which there will be a main “export” button on the homepage which will export a csv for the Trader to upload to his trading management software.

7) Git Repository (0 points): Create a git repository and share it with Github or Bitbucket (handle:- pkarthik88)

* [https://github.com/cheewengg/it5007-project/](https://github.com/cheewengg/it5007-project/tree/report)
  + Branch Code is for Chee Weng’s work.
  + Branch Code-Python is for Gary’s data pre-processing work.
  + Branch Report is for any general documentation such as this document.