|  |
| --- |
| Description: 18897 CIC A4 Portrait WordTemp_cropped.jpg |
|  | |
|  | |
| |  |  | | --- | --- | | **SUBJECT NUMBER & NAME** | 36102 iLab1 | | **NAME OF STUDENT**  **(PRINT CLEARLY - SURNAME, FIRST NAME)** | LEUVA UJASHKUMAR | | **STUDENT ID NUMBER** | 12739975 | | **STUDENT EMAIL** | [Ujashkumar.leuva@student.uts.edu.au](mailto:Ujashkumar.leuva@student.uts.edu.au) | | **STUDENT CONTACT NUMBER** | 0470331919 | | **DUE DATE** | 12/08/18 | | **ASSESSMENT ITEM NUMBER/TITLE** | AT1- Project Design Journal  Part A : Learning Goals and Weekly Entries | | * I confirm that the work submitted conforms with the university’s guidelines on academic integrity.   *Refer to the UTS policy on ‘Advice to Students on Good Academic Practice’*: <http://www.gsu.uts.edu.au/policies/academicpractice.html>   * I am aware of the penalties for plagiarism. This assignment is my own work and I have not handed in this assignment (either part or completely) for assessment in another subject. * If this assignment is submitted after the due date I understand that it will incur a penalty for lateness unless I have previously had an extension of time approved and have attached the written confirmation of this extension.   Please provide details of extensions granted here if applicable \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **Signature of Student:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **­­Date:** 12/08 /2018  If submitted electronically tick here to indicate you agree with the above | | | |

**Introduction:**

Our Project is based on to develop stock earning driver model and predicting the stock market for next year from analyzing the data from last 10 to 20 years.

**Background and Project goals**

Our project is about to develop stock earning driver model for J.P. Morgan. Our primary project goal is to identify the key drivers of individual stocks earnings, such as Iron Ore prices and FXRates for FMG-AU, Oil prices, interest rates, consumer sentiment for QAN-AU, Oil+ Coal+ LNG+ wholesale energy prices + temperature + solar panel take-up rates for ORG-AU using Python and Google Cloud to predict the stock market for next year for J.P. Morgan.

**Status report and Design journal**

As our project is to predicting the stock market, our team is occupied in the collection of datasets and review the relationship between the data.

Our team has to collect the datasets from a different data source about the stock market as well as the foreign exchange rate from past 10 to 20 years, as well as create a grouping between those datasets. Our team also have to develop a draft of objectives, expected outcomes, and project delivery. Our team had already collected historical data about oil, mining, forex data as well top 200 organization by revenue in Australia. From next week, our team will started working on data cleaning and manipulation. Our team will follow CRISP-DM methodology for the implementation of this project.

In this Automated model to we have to identify and predict 12 month earnings of companies based on these ‘key drivers’, including logical explanations (perhaps with charts, NLG text etc).

**The project design journal**

| **Assessment Criteria** | **Part A: Identified elements** |
| --- | --- |
| Evaluated data challenges and selected appropriate approaches to data discovery | **Evaluated Data Challenges**   * In this dataset there is enormous amount of data which is about more than ten million rows as there is daily stock data from past 20 years. * Collecting a dataset from different data sources are herculean task for both of us as well as the data scraping for sentiment analysis- twitter data and news data.   **Approaches to Data discovery**  For the data collection, I tried a Bloomberg before but I decided to look up on open data sources and API to figure out the datasets. Yahoo Finance is one API which seems to have the majority of the data-sets apart from iron ore. |
| Identified key concepts, frameworks or processes to utilize for problem solving | **Identified Key concepts and frameworks**  I broadly did research on two key data-frameworks around finance in my first three weeks and I intend to use these methods when I model on the data sets  **Hierarchical Cluster Portfolio**- hierarchical clustering on the covariance matrix of stock returns and then find a diversified weighting by distributing capital equally to each cluster hierarchy (so that many correlated strategies will receive the same total allocation as a single uncorrelated one).  **Random Matrix Filtering**– Random matrix theory (RMT) filters, applied to covariance matrices of financial returns, have recently been shown to offer improvements to the optimisation of stock portfolios.  Random Matrix Filtering is a recently introduced filtering procedure, and demonstrate the applicability of this method in a controlled, simulation environment. |
| Considered and applied legislation and standards for managing data in stakeholders’ context | Due to some privacy issues, J.P. Morgan denied to provide dataset outside firm, so myself decide to collect the datasets from open source. |
| Explored effective Data visualization techniques (**Self-developed criteria**) | I intend to identify visualization techniques to represent Hierarchical Cluster Portfolio and Random Matrix Filtering |
| Project management skills **(Self-developed criteria)** | I have set-up a github channel to Keep track of tasks, timelines with well defined deadlines. |

**Learning prospective for this project**

As I already experienced in Python along with data visualization libraries, Myself do not have to spend much more time on it. However, I hadn’t applied machine learning algorithms as well as myself is entirely new into sentiment analysis, hence before I started working on it, I should test the waters. In particular, Finance and banking domain is new for me. Thus myself have to understand in a much broader context as well as continue reading about the stock market.