**Case Study 1 – Understanding Consumer loans**

Note: this case study is based on a proprietary research and should be only used for the case study

**Case Study Scope:**

* The focus of this case study are consumers in UK.
* We do not expect you to have access to any private or paid data sources or tools.
* Please note that the data is to be used for the purpose of the case study only.
* The information given is to help you get some context and is not necessarily representative or exhaustive.

**Your tasks:**

1. Factors that determine Consumer loans performance

* Find suitable visualizations to generate new insights from the provided data. Briefly describe your key insights that became apparent in producing these visualizations.
* What are the factors you think should be considered to understand and predict consumer defaults?
* Come up with a suitable quantitative framework to understand the Key drivers defaults and predict their probability in a 12 and 24 month period.

1. Dataset

* Contains information of 72,167 different consumer loans started in the period 2014-18. You will find three tabs in the dataset.
  + Static Data: Contains top level data for each loan
  + Performance: Contains the month by month evolution of each loan
  + Dictionary: A data dictionary of the variables.

1. Assessment criteria

* You will be assessed on the approach you use.
* You are free to choose any form of presentation you are comfortable with.
* Be prepared to give a short overview on the tools and programming code you utilized.
* You are strongly encouraged to be creative. Try to find additional data to enhance your forecast and to model indicators that create insights about the consumer defaults.

**Case Study 2 – Analysing earning calls and its consequences in bond performance**

Note: this case study is based on a proprietary research and should be only used for the case study

**Case Study Scope:**

* The focus of this case study are listed companies around the world
* We do not expect you to have access to any private or paid data sources or tools.
* Please note that the data is to be used for the purpose of the case study only.
* The information given is to help you get some context and is not necessarily representative or exhaustive.
* Earning calls are regular discussions/presentations that company management has with investors regarding past performance and prospects.

**Your tasks:**

1. Factors that determine bond performance

* Find suitable visualizations to generate new insights from the provided data. Briefly describe your key insights that became apparent in producing these visualizations.
* Extract data and metadata from the earning calls transcripts
* Come up with a suitable quantitative framework to understand the Key drivers behind bond spread movements
* How to better translate language into insightful data?
* Explain why your variables should be associated with bond performance
* What are the factors you think should be considered to understand and predict bond performance. What other datasets should be added to this analysis?

1. Dataset

* The datasets are large, so you might select to work only on a sample. There are two datasets for this case study:
  + BondSpreads(CSV File): Contains the daily spread information 7,751 different bonds
  + FixedIncomeCalls(PDF files): Are the transcripts of each call. On the header you will find company metadata

<https://drive.google.com/open?id=1qcCSPbmJ00LzdDa3VlvyyVLgxj4w1fiq>

* Note(1): Each company as a unique identifier (Company Ticker/Equity ticker) that can be used to link both datasets
* Note(2): A company can have more than one bond
* Note(3): A company can have more than one earning call
* Note (4): Bonds have different life cycle. So for a given date a bond might simply not exist.

1. Assessment criteria

* You will be assessed on the approach you use.
* You are free to choose any form of presentation you are comfortable with.
* Be prepared to give a short overview on the tools and programming code you utilized.
* You are strongly encouraged to be creative. Try to find additional data to enhance your forecast and to model indicators that create insights about the consumer defaults.

**Case Study 3 – Location Based Intelligence**

**Case Study Scope:**

* The focus of this case study restricted to finding good locations for Real Estate Investment in **Glasgow**
* The key factor that will be important is that the location should either be currently undervalued/ display strong potential to appreciate in the next 3-5 years.
* We do not expect you to have access to any private or paid data sources or tools
* You are free to use any publicly available data sources

**Your tasks:**

1. Find factors that are important for Real Estate Prices:

* Find suitable data that might be useful to understand real estate prices
* We are not looking precision but rather the though process and creativity
* Using samples of data rather than large datasets to demonstrate the concept would suffice
* Find suitable visualizations to highlight insights and help communicate results to business users
* Optional:
  + Build machine learning/ deep learning models to predict location based prices
  + Think of ways this can be made interactive for the business user
  + Develop an approach to expand this using paid datasets/ tools/ better techniques etc.

1. Dataset

Hint:

1. Analyse this dataset - <https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads>
2. Social media often contains location information that can be utilized here
3. Assessment criteria

* You will be assessed on the creativity of the approach you use.
* You are free to choose any form of presentation you are comfortable with. A working demo/ application would be a bonus
* Be prepared to give a short overview on the tools and programming code you utilized.

**Case Study 4 – Handling Large Unstructured Datasets**

**Case Study Scope:**

* The focus of this case study is to utilize twitter data to find consumer patterns/ economic trends
* We would like to use the twitter data to understand the need for financial products in the UK market
* We do not expect you to have access to any private or paid data sources or tools
* You are free to use any other publicly available data sources

**Your tasks:**

1. Capture tweets and summarize them to provide useful info:

* Capture, store and summarize twitter data
* We are not looking precision but rather the though process and creativity
* Using samples of data rather than large datasets to demonstrate the concept would suffice
* Find suitable visualizations to highlight insights and help communicate results to business users
* Optional:
  + Think of ways this can be made interactive for the business user
  + Develop an approach to expand this using paid datasets/ tools/ better techniques etc.
  + Utilize machine learning/ AI models for text/ time series prediction

1. Dataset

Hint:

1. You may build your own scraper to capture twitter streams in real-time or use an archive such as - <https://archive.org/details/twitterstream>
2. Assessment criteria

* You will be assessed on the creativity of the approach you use rather than the accuracy
* You are free to choose any form of presentation you are comfortable with. A working demo/ application would be a bonus
* Be prepared to give a short overview on the tools and programming code you utilized.