

1. Your group needs to decide what kind of project you are going to work on and lock in your decision.

"Predictive modelling for credit card fraud detection." A finance based project where we investigate whether new credit card transactions are fraudulent or genuine based off of a simulated database.

2. You need to submit a free style paper that describes your project on a high level. Please cover the following questions:

1. What kind of data research and analysis are you going to take on?

- a. Machine learning model
- b. Analysis prediction
- c. API

2. What industry or areas does it cover?

- a. Banking
- b. Finance
- c. Insurance
- d. Credit card fraud
- e. Crime

3. What questions are you planning to answer?

- a. **Are new transactions fraudulent or not? - Focus Question of Project**
- b. How can we segment and cluster fraudulent transactions? - *Sub Question related to Focus Question*
- c. What are the crime rates in comparison to the fraud rates? - *Sub Question related to Focus Question*

4. What data sources are you planning to use?

- a. Credit Card Transactions Fraud Detection Dataset :
<https://www.kaggle.com/datasets/kartik2112/fraud-detection?resource=download&select=fraudTrain.csv> Simulator by Brandon Harris

Harvard Referencing of Simulated Dataset: Shenoy, K. 2020. Credit Card Transactions Fraud Detection Dataset. Kaggle. [Online]. [Accessed 28 April 2022]. Available from: <https://www.kaggle.com/datasets/kartik2112/fraud-detection>.

- b. For Simulator see: Harris, B. 2020. Sparkov_Data_Generation. GitHub. [Online]. [Accessed 28 April 2022]. Available from: https://github.com/namebrandon/Sparkov_Data_Generation.
- c. Provisional API resources list:
Federal Bureau of Investigation. Crime Data Explorer. [Online]. [Accessed 06 May 2022]. Available from: <https://crime-data-explorer.fr.cloud.gov/pages/docApi>.

5. Describe the team approach to the project work: how are you planning to distribute the workload, how are you managing your code, how are you planning to work on your project?

- a. Team approach will be based on Agile Methodology
- b. Have quick 15 min calls post class every day to discuss where we are at and what needs to be done, similar to a daily stand up meeting.
- c. Possibly extend one evening call to be longer (45-60 mins) to be more detailed.
- d. Work through the roadmap of the report together {this will also solidify what the tasks are for our project/report}

- e. Regularly employ the following platforms to share with each other progress and resources relevant to the project:
 - i. Slack
 - ii. Google Drive
 - iii. Google Collab
 - iv. GitHub
 - v. Jupyter Notebook
- f. Delegate graphs to be done by individuals or one individual to do all graphs
- g. Make sure to note down tools and libraries used as well as what references, datasets, databases, APIs etc.

Task Division

Modelling Taskforce (2 people) - Juliana & Ilaria

Research best classification models for fraud prevention, split dataset into train and test, fit the model into the data, analyse results (at least two different models).

API & Visualisation Taskforce (2 people) - Nicol & Joey

Find an open access API that could complement our analysis on credit card fraud (example: crime, weather). Request data from the API and communicate with the visualisation taskforce to integrate it into visualisation. Perform meaningful exploratory analysis with data, using data to tell a story and convey an idea. (Example: plot map of credit card fraud, show main institutions involved, etc).

Reporting, sliding and project management Taskforce (1 person) - Rumaanah

Make sure everyone is doing their tasks, help colleagues if anyone has any issue, keep track of the work, create the reports. Make sure the parts of the project fit together. Point of contact in the communications with Polly (instructor).

Project Process Flowchart:

