

## List of Projects

This is a list of projects that I have worked on as part of my graduate data science course.

### 1. Historical Sales and Active Inventory

*This project is about building a binary classifier which gives a list of product IDs which need to be retained in the inventory or list of products that need to be removed using Exploratory Data Analysis.*

*Link to the repository:* [https://github.com/samantoz/BU\\_DSC530\\_TermProject](https://github.com/samantoz/BU_DSC530_TermProject)

### 2. Stock pick based on market data

*This project is to answer the question: Is it possible to build a dataset for a machine learning model that could learn the nature of a stock based on the company performance and then leverage this knowledge in order to predict which stock will be worth buying?*

*Link to the repository:* [https://github.com/samantoz/bu-datascience/tree/master/DSC540-DW/Term\\_Project](https://github.com/samantoz/bu-datascience/tree/master/DSC540-DW/Term_Project)

### 3. Predicting Heart disease using regression analysis

*This project is to do EDA on the data to get an understanding of the different features and how they are related to one another. We are trying to build a classification problem to answer our main question about the chances of having a heart disease.*

*Link to the repository:* <https://github.com/samantoz/bu-datascience/tree/master/Statistical%20Analysis%20with%20R>

### 4. Weather prediction based on zip code

*This is a program written in python that interacts with the webservice in order to obtain weather data. It will prompt the user for their city or zip code and request weather forecast data from OpenWeatherMap. This program will display the weather information in a READABLE format to the user.*

*Link to the repository:* [https://github.com/samantoz/bu-datascience/tree/master/dsc\\_510](https://github.com/samantoz/bu-datascience/tree/master/dsc_510)

### 5. Dashboard: Is Airline Travel more riskier

*This is a dashboard that shows the Airlines Fatalities trend over the years for a fictitious airlines company and hence tries to send out a social media campaign using data to prove that airlines travel is still the safest.*

*Link to the repository:* [https://github.com/samantoz/dsc640-data\\_presentation/tree/main/term\\_project](https://github.com/samantoz/dsc640-data_presentation/tree/main/term_project)

### 6. Context Aware Text Analysis

This is the Enron email text analysis. I am interested to classify(topic classification) the emails that was released for general public.

*Link to repository:* <https://github.com/samantoz/dsc680-applied-datascience/tree/main/Project-1>

## **7. Classification of news category based on the headline**

*This is a very small text analysis project that tries to categorize the news items into predefined categories based on the headlines.*

*Link to repository:*

[https://github.com/samantoz/bu\\_dsc/blob/master/notebooks/DSC550/submitted/project\\_article\\_classification.ipynb](https://github.com/samantoz/bu_dsc/blob/master/notebooks/DSC550/submitted/project_article_classification.ipynb)

## **8. Home price prediction**

*This is a project that we worked on as a team. Our object is to discuss major factors which influence the housing price and make precise predictions for it. The model provides price prediction of similar comparable of people's dream homes, such that the buyers and sellers can better negotiate the price as per the market trend.*

*Link to the repository:* [https://github.com/samantoz/bu\\_dsc/tree/master/notebooks/DSC630](https://github.com/samantoz/bu_dsc/tree/master/notebooks/DSC630)

## **9. Yet to plan**

## **10. Yet to plan**