

ST 1 Assignment 9 **Capstone Programming** **Project**

Project submission due date: Week 13 Sunday 23:59

This assignment is worth 30 marks

(Project Report: 20 marks)

(video presentation recording and Code review

Interview: 10 marks)

Project Description and Requirements

Theme: **Data Analytics, Visualization, Prediction and Deployment**

Objectives: Demonstrate the understanding of some of the following Python Libraries/Packages: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn, PyTorch, Keras, TkInter, Flask, Django or any similar open-source Python packages approved by tutor/unit convenor.

Group/Solo Participation Requirements:

- i. Students can work in groups, with following requirements:
 - a. This assignment can be done in a group of 2 students or as a solo project.
 - b. The project tasks and requirements will be same whether you work solo or in a group (As best practice approach in Software Engineering field is to work on group/team based coding activities. Pair Programming: https://en.wikipedia.org/wiki/Pair_programming)
- ii. The tutorial/lab activities in week 10, week 11, and week 12, will provide further guidance on sample project tasks using (Pycharm, Google Colab and Replit IDEs) and with appropriate submission templates for reports and presentations for this capstone project activity.

Capstone Task:

I. Dataset Description

- A links for each dataset option (numbered 1 to 20) for working on the Capstone Project task is given in Appendix I.
- You need to select the dataset from the following allocation table according to the tutorial group you are in.

Tutorial Group	Dataset Options (Choose One)
4483 C/01- Wednesday 16.30- 18.30	1,2
4483 C/02- Thursday 13.30- 15.30	3,4
4483 C/03- Thursday 14.30- 16.30	5,6
4483 C/04- Wednesday 14.30- 16.30	7,8
4483 C/05- Wednesday 9.30- 11.30	9,10
4483 C/06- Thursday 9.30- 11.30	11,12
4483 C/07- Tuesday 13.30- 15.30	13,14
4483 C/08- Wednesday 15.30- 17.30	15,16
4483 C/09- Wednesday 17.30- 19.30	17,18
8995 C/06- Thursday 11.30- 13.30	19, 20

II. Dataset Analysis and Prediction Studies:

Steps required for completing the Data Analysis and Prediction Studies, for the dataset you have selected from the allocation table is as outlined below:

1. Reading the dataset
2. Problem statement definition
3. Target variable identification
4. Visualising the distribution of Target variable
5. Data exploration at basic level
6. Identifying and Rejecting useless columns
7. Visual Exploratory Data Analysis of data (with Histogram and Barcharts)
8. Feature Selection based on data distribution
9. Removal of outliers and missing values
10. Visual and Statistic Correlation analysis for selection of best features
11. Data Conversion to numeric values for machine learning/predictive analysis
12. Training/Testing Sampling and K-fold cross validation
13. Investigating multiple Regression algorithms
14. Selection of the best Model
15. Deployment of the best model in production

III. Submit the final project report in the canvas and create one **GitHub repository** for your Capstone project and include the link in the report.

IV. Prepare a **15-minute presentation** about your capstone project findings and submit the presentation slides and the **Video Recording** of your presentation.

- V. Give an 5-10 minute **code review interview about selected parts of the project** chosen by the unit tutor in **week 13 in your tutorial time-slot**.

Appendix 1: Dataset Options

1. <https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction>
2. <https://www.kaggle.com/datasets/deepcontractor/car-price-prediction-challenge>
3. <https://www.kaggle.com/datasets/mrsimple07/laptoppriceprediction>
4. <https://www.kaggle.com/code/karnikakapoor/diamond-price-prediction>
5. <https://www.kaggle.com/datasets/avikasliwal/used-cars-price-prediction>
6. <https://www.kaggle.com/datasets/mrsimple07/clothes-price-prediction>
7. <https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction>
8. <https://www.kaggle.com/datasets/mohannapd/mobile-price-prediction>
9. <https://www.kaggle.com/datasets/shawkyelgendy/furniture-price-prediction>
10. <https://www.kaggle.com/datasets/sid321axn/gold-price-prediction-dataset>
11. <https://www.kaggle.com/datasets/harishkumardatalab/medical-insurance-price-prediction>
12. <https://www.kaggle.com/datasets/mexwell/boat-price-prediction>
13. <https://www.kaggle.com/datasets/stevezhenghp/airbnb-price-prediction>
14. <https://www.kaggle.com/datasets/amohankumar/tesla-stock-price-prediction-dataset>

15. <https://www.kaggle.com/datasets/ialabISEP/footballsoccer-statistics>
16. <https://www.kaggle.com/datasets/nelgiriyeWithana/australian-vehicle-prices>
17. <https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction>
18. <https://www.kaggle.com/datasets/theDevastator/football-price-prediction>
19. <https://www.kaggle.com/datasets/olegshpagin/starbucks-stock-price-prediction-dataset>
20. <https://www.kaggle.com/datasets/bhuwanesh340/ecommerce-price-prediction/data>

*******End of Capstone Project Description *******

ST1(4483) Assignment Capstone Project Marking Rubric

Capstone Project submission due date: Week 12 Sunday 23.59pm

This assignment is worth 30 marks (02+10)

Student Name/ID: Student Name/ID:		Unit: 4483/8995	Group/Solo		
Project/Dataset Title:					
Project Deliverable ID	Project Task	Maximum Marks	Marks Obtained	Comments	
1.	Reading the dataset	0.5			
2.	Problem statement definition	0.5			
3.	Target variable identification	0.5			
4.	Visualising the distribution of Target variable	1			
5.	Data exploration at basic level	1			
6.	Identifying and Rejecting unwanted columns	0.5			
7.	Visual Exploratory Data Analysis of data (with Histogram and Barcharts	1			
8.	Feature Selection based on data distribution	1			
9.	Removal of outliers and missing values	1			
10.	Visual and Statistic Correlation analysis for selection of best features	1			
11.	Data Conversion to numeric values for machine learning/predictive analysis	1			

12.	Training/Testing Sampling and K-fold cross validation	1		
13.	Investigating multiple Regression algorithms	2		
14.	Selection of the best Model	1		
15.	Deployment of the best model in production	2		
16.	GUI/WEB Deployment using either tkinter/Flask/Streamlit	5		
17.	Presentation Slides and Video Recording	5		
18.	Code Review Interview	5		
Total Marks		30		
Further Comments				

