

a. What is the task about?

This project required us to dive into many mathematical and statistical topics, starting from analysing different datasets with specific scenarios. We were asked to understand hypotheses and evaluate code to obtain desired results. Our goal was to present our thought process throughout the assignment by adding code markdowns and comments. Every step taken in the assignment was regarding the bigger task and story that we were supposed to solve.

b. What Maths and Statistics are involved?

Exploratory analysis techniques: plots, charts, basic exploratory dataset functions.

Inferential analysis techniques: ANOVA, correlation, Chi-square.

Predictive analysis techniques: regression, classification, decision trees.

Loss functions: L1 loss, L2 loss, Log loss, Categorical cross-entropy loss, Hinge Loss.

Classification and Regression metrics: RMSE, R-squared, Accuracy, Precision, Recall, F1 Score, Confusion Matrix.

c. How was the implementation done?

During the assignment, I used many libraries, mostly for math, analytics, and plots. My goal was to create clean code that is easy to understand with necessary comments. I used algorithms that fulfilled the criteria of the task, primarily relying on Stack Overflow for debugging and other websites for algorithmic code. In my assignment, I tried to use as much as possible the code that I had created previously to achieve a better workflow.

d. What were the outcomes of each sub-task?

The outcomes of sub-tasks were mostly some metrics or plots that I interpreted and created implications from. Every sub-task had to be approached regarding the goal we wanted to achieve, and the approach should be adjusted to the dataset we are operating on. Analysing the dataset, starting from the exploratory part to more complex calculations, was often the next step for further evaluation.

e. What challenges did you face, and how were they resolved?

Most challenges involved difficult metrics to apply in our thought process, but there are good materials from our classes and the internet to operate with those terms more fluently. Although some tasks required more time, the obstacles could be resolved with the proper approach. All in all, with a good understanding of concepts and theory, proper result validation was given to the task.

f. References

These are references I used in during creating my project:

class lectures, <https://statisticsbyjim.com/anova/f-tests-anova/>,
<https://www.statology.org/a-simple-guide-to-understanding-the-f-test-of-overall-significance-in-regression/>,
<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.KernelPCA.html>,
<https://www.geeksforgeeks.org/linear-mapping/>,
<https://brilliant.org/wiki/kernel/#:~:text=The%20kernel%20%28or%20nullspace%29%20of%20a%20linear%20transformation,R%2D%5En%20Rn%20whose%20dimension%20is%20called%20the%20nullity>,
<https://www.geeksforgeeks.org/decision-tree/>