Reflective Diary Template

Group Number: 113

Full Name	Pradnya Rajendra Alchetti
Role in group:	Developer
Role responsibilities:	Fixing errors in dirty_data.csv ● Detecting and Fixing date format ● Detecting and Fixing time plus order_type criteria ● Detecting and Fixing branch code based on order_id ● Detecting and Fixing customer_lat and customer_lon ● Detecting and Fixing distance_to_customer_KM Imputing missing values: ● detecting and imputing branch_code and distance_to_customer_KM Detecting Outliers: ● EDA for detecting outliers
Contribution to Group:	50%
Learnings from Project	Learned how to detect errors using exploratory data analysis such as summary, graphs, etc. Learned to fix the errors and impute the missing values with proper suitable values. Learned how to detect outliers through EDA.
Learning Techniques	The documentation of networkx to determine the shortest path between nodes using Dijkstra's algorithm helped me to understand how the shortest path between two locations (given latitude and longitude) can be determined. The lecture slides helped me to understand how the residuals can be used to detect outliers in the given dataset
What went well	The group member's contribution to the assignment.
What went wrong?	We started the assignment a bit late so found it difficult to explore the entire data
Resolution to solve issues:	We will be starting to work on the assignment as soon as it is released
Overall Conclusion: The given assignment had several challenging tasks such as determining the shortest path using Dijkstra's algorithm, using linear equations to determine order_price and using linear regression to determine the delivery fee. This assignment helped us to learn how to perform EDA to detect errors and fix those errors.	

How would you do it, if asked to do it again? I would like to first perform the exploratory analysis on the entire data and then try to fix the errors.

Full Name	Akshatha Shivashankar Chindalur
Role in group:	Developer
Role responsibilities:	Task 1: Missing Data 1. Constructed linear models for the 3 branches 2. Imputed the deliver fee column of the dataset Task 2: Dirty Data 1. Derived and solved simultaneous equations to obtain prices of items 2. Identified the unique items in each meal 3. Provided a suitable fix for order_type and order_price 4. Identified and fixed errors in customerHasloyalty? Task 3: Outliers Detection 1. Computed the residuals 2. Detected and removed outliers Documented and cleaned up the code
Contribution to Group:	50%
Learnings from Project	1. The assignment helped to understand the different errors (syntactic, semantic and coverage) that could occur in dirty data. 2. Learnt how regression can be used for imputation and outlier detection.
Learning Techniques	1. Online documentation and blogs about outlier detection and missing value imputation. 2. Trying tutorial materials and reading lecture notes.
What went well	Team effort
What went wrong?	Due to busy schedule was unable to perform extensive EDA before fixing errors.
Resolution to solve issues:	Time management could be better.

Overall Conclusion: The assignment was challenging and required contribution from both the members equally. It also helped to get an understanding about the different phases in data wrangling. Also, learnt the need for time management.

How would you do it, if asked to do it again? Explore the data with more plots and figures.