Problem Description

You (who are working as a consultant) have been approached by the senior product manager for a major app-based cab company (Ola, Uber, and Meru are examples of app-based cab companies). The product manager tells you about the major metrics he reviews, week-on-week:

- **Demand**: Number of requests made by customers
- **Fulfilment** %: Out of the requests made, for what % of requests was a cab allocated
- **Utilization** %: Amount of time cars are occupied with a customer as a fraction of the total number of hours all cars have logged
- **Cancellation** %: The number of requests cancelled is further split as 'by customer' and 'by driver'

While going through these numbers he realizes that while all numbers seem to be healthy for the city of Hyderabad, the cancellation % seems to be higher. To check and validate that he compares it to the national average and against the rest of the top cities and realizes that it is a full 6% higher in Hyderabad. He is confused about this anomaly.

Was there something that happened this week with the city? Are drivers cancelling more? Did their pricing algorithms fail and caused prices to increase and thus more customers are cancelling? Based on following questions analyze your data set and identify the reason behind cab cancellations in Hyderabad.

Q. Using a bar chart, compare the average total cancellations by city. Does Hyderabad have maximum cancellations percentage?

Q. Using a side-by-side bar chart, compare average customer cancellations v/s driver cancellations across various zones in Hyderabad. Note your observations

Q. Using line chart, Compare total cancellations, customer cancellations and driver cancellations over days. Note your observations

Q. Using stacked-bar chart, observe customer cancellations across zones and 4 hour window. Are cancellations confined to any particular time window or zone?

Q. Using scatter charts understand which factors have a strong correlation with cancellation. Is it the price, the demand, the ETA etc.?