Problem statement

You (are working as a consultant) have been approached by the senior product manager for a major app-based cab company (Ola, Uber, 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 %: Number of requests cancelled 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, cancellation % seems to be higher.

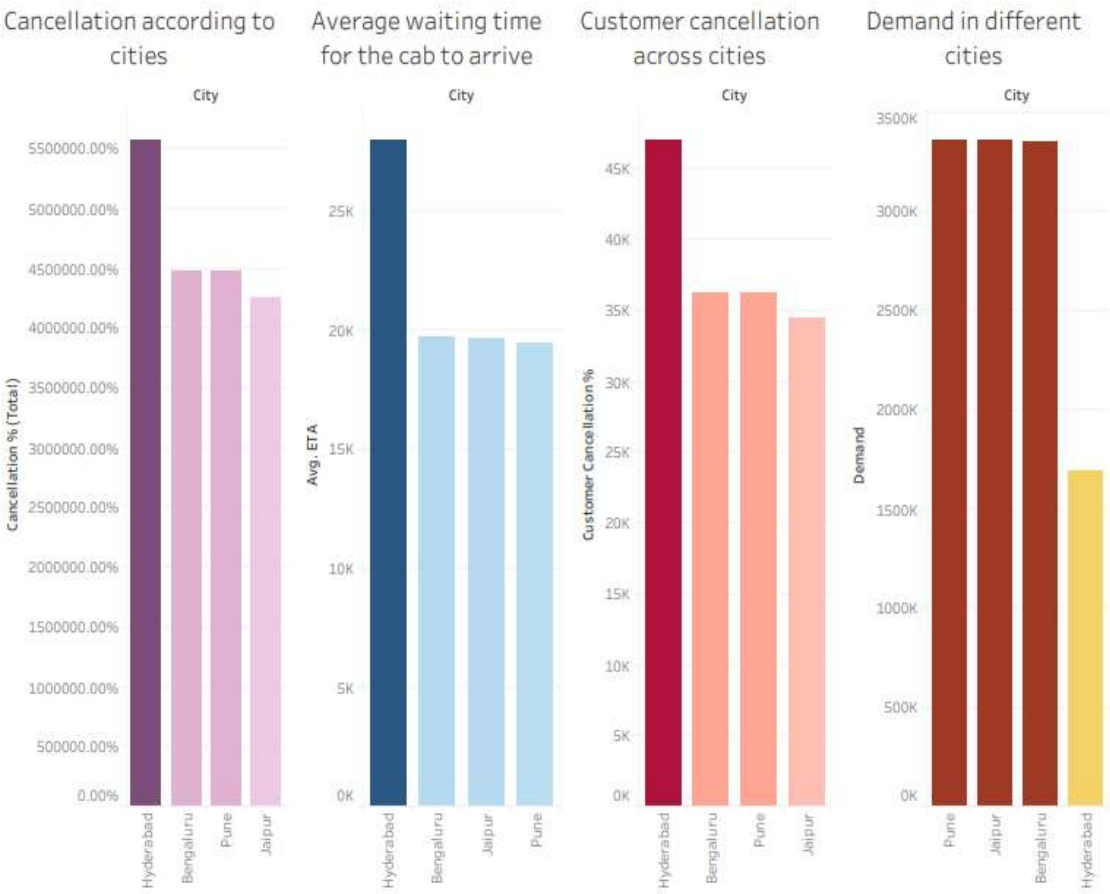
To check and validate that he compares it to the national average and against the rest of the top cities and realizes that it is a full 6% higher in Hyderabad. He is confused with this anomaly. Was there something that happened this week with the city? Are drivers cancelling more? Did their pricing algorithms fail and caused prices to increase and thus more customers are cancelling?

Q. You want to understand whether the problem of cancellation is confined to Hyderabad only, whether it's confined to a particular zone in Hyderabad and whether it's more prevalent for a specific period of 4 hours for a day.

Q. You also want to understand which factors have a strong correlation with cancellation. Is it the price, the demand, the ETA etc.? Import the Dataset to Tableau and based on the given hints analyse the data. Create the charts using Tableau and paste them as screenshot in a word file and submit the word file. Add your insights about the data in the word file along with the screenshots

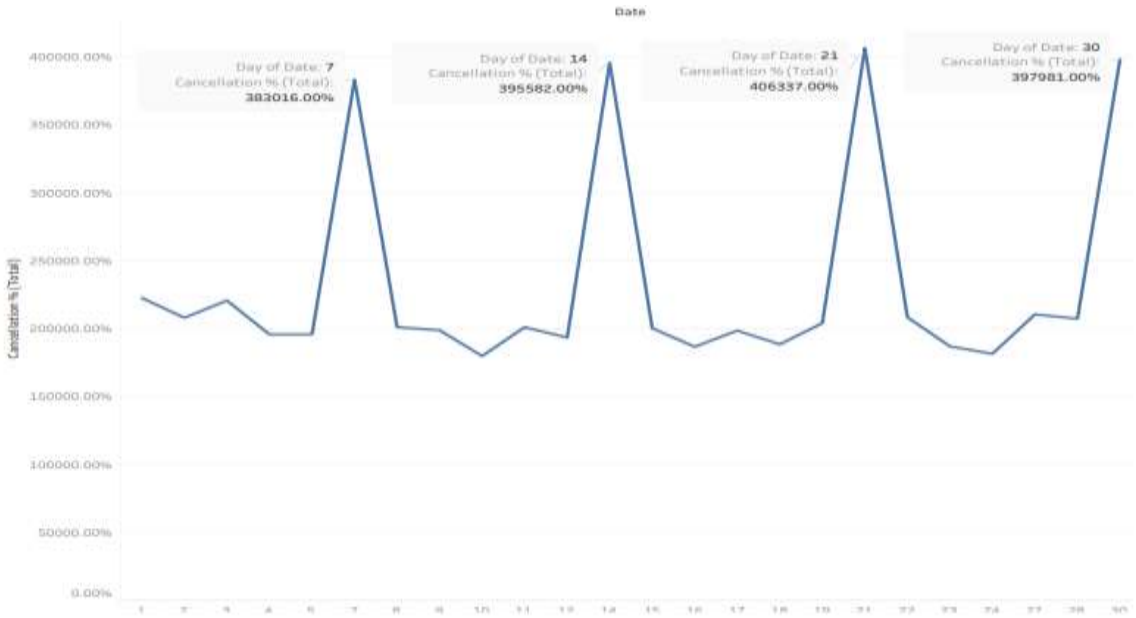
Solution

After comparing data among different cities we have observed that there has been highest cancellation of cabs across Hyderabad city and there can be several factors behind it such as demand across cities, customer cancellation factors, Avg ETA, Avg trip size, 4 hour window etc. So after plotting it in tableau we have observed that Avg ETA i.e. the waiting time for the cab to arrive at the desired location, customer cancellation has been the highest in Hyderabad city and the demand for the cabs is lowest in the Hyderabad city.



We have also compared several other factors across cities such as cancellations according to 4 hour window in Hyderabad, driver cancellations across cities, average trip size and average price across the cities and the fulfilment % across the cities but they are not much of significance or the contributing factor for the cab cancellation in Hyderabad.

Further if we check the days in month we will come to know about the interesting observation i.e. Day7, Day14, Day21 and Day30 have the maximum number of cancellations.



By comparing all the data points across different cities and different zones we have come to the conclusion that Average ETA i.e. average waiting time for a cab to arrive in the give time slot for a given zone is the single most important factor which is responsible for the surge in cancellation of cabs across the Hyderabad city.

City	Avg. ETA	Avg. Price	Avg. Trip Size	Cancellation % (Total)	Customer Cancellation %	Demand	Driver Cancellation%
Pune	19,510	261,659	19,400	4486697.00%	36,239	3,361,181	8,628
Bengaluru	19,722	261,871	19,238	4488636.00%	36,254	3,358,281	8,632
Jaipur	19,653	261,793	17,523	4265120.00%	34,449	3,360,820	8,202
Hyderabad	28,009	256,992	16,842	5572658.00%	47,019	1,690,404	8,707