

**BI Project**  
**Affan Ali Khan**  
**18649**

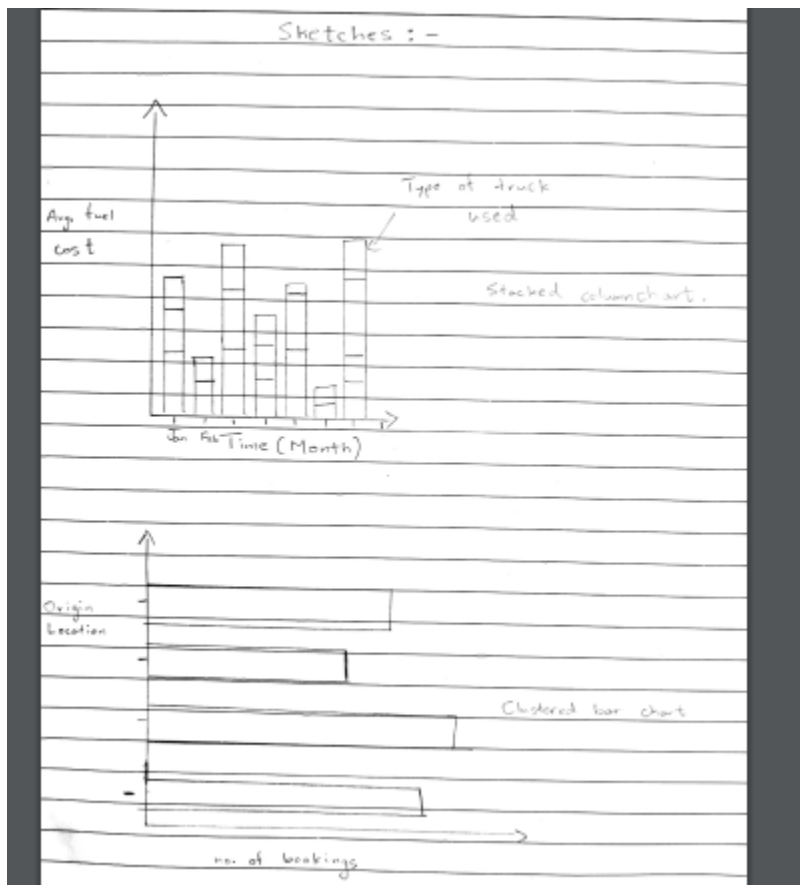
Dataset:

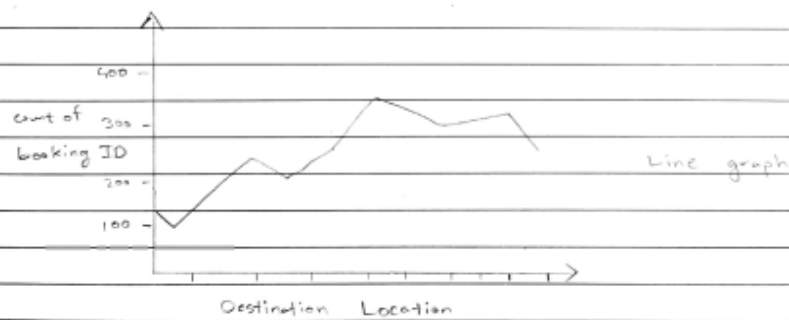
The dataset I have chosen is the delivery trucks dataset.

<https://www.kaggle.com/ramakrishnanthiyagu/delivery-truck-trips-data>

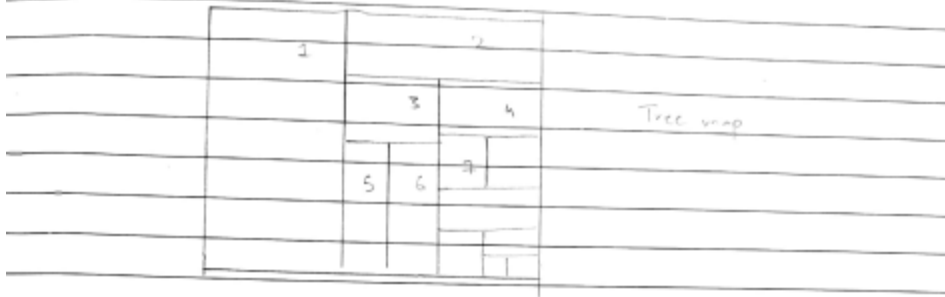
Sketches:

Sir I deeply apologise, I know that the sketches are a bit hard to read, but they were coming out this way in the CamScanner app. I hope you can understand them. Thankyou so much.





Deliveries by GPS Provider



		s	k	Tree wrap
--	--	---	---	-----------

Supplier and successful deliveries.

100%



1000	Supplier 2
------	------------

	1000	Supplier 2
--	------	------------

800	Supp 2
-----	--------

800 Supp 2

700	Supp 3	
-----	--------	--

700	Supp 3	
-----	--------	--

700 Supp 3 Funnel chart

680

Supp. 4

680 Supp. 4

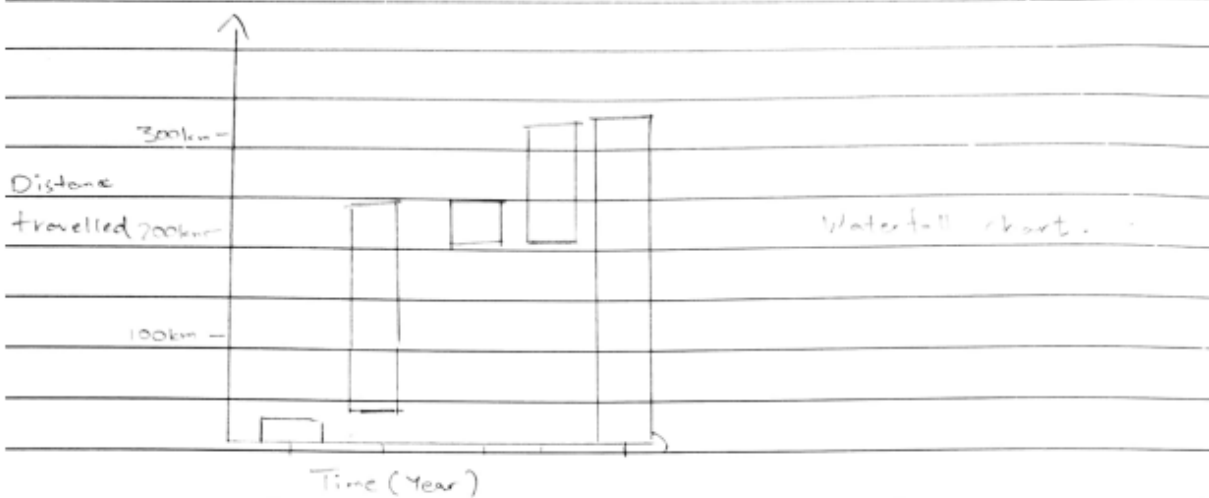
$$\boxed{500} \quad S_{\text{sup}} 5$$

$$\boxed{500} \quad S_{up} \ 5$$

350 Supp 6

350 Supply

35%

[illegible]

## Tool:

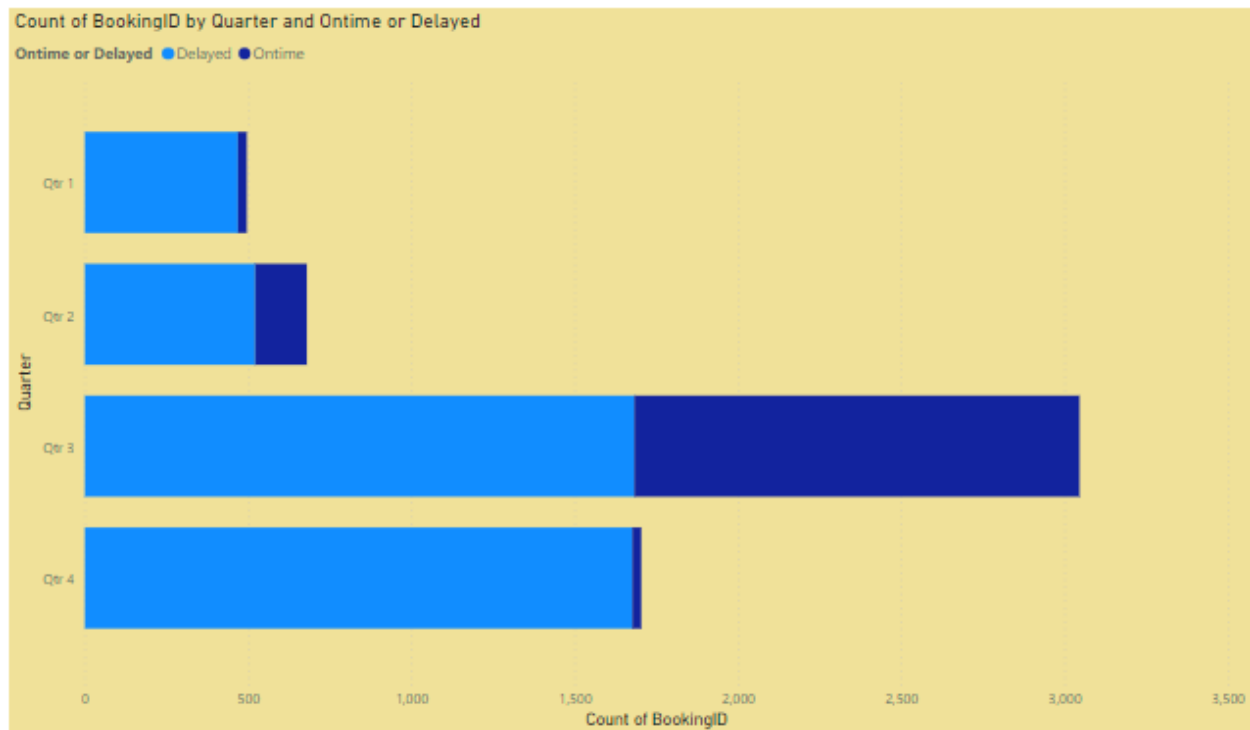
The tool I have used for wrangling and analysis is PowerBI.

## Wrangling and Cleaning:

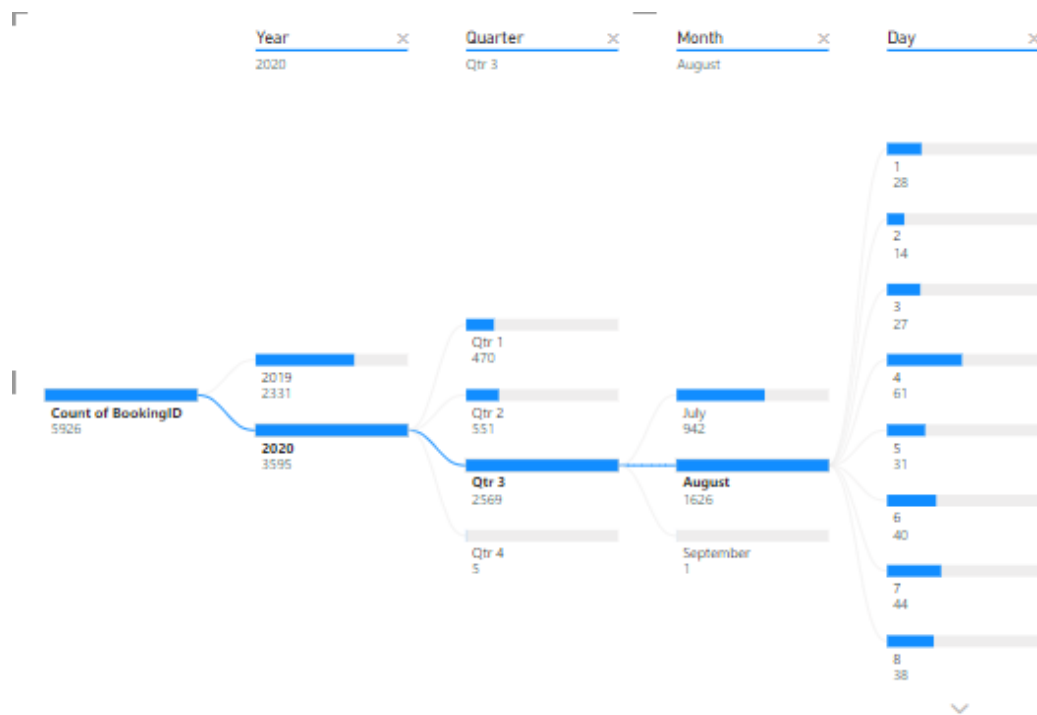
- There were many rows which contained null values in many of the columns in that row so I removed or dropped all such rows.
- In the TRANSPORTATION\_DISTANCE column there were many null values so I used the average of the other values in the column to fill the null values.
- There were two outlier values in the Planned\_ETA column, with year 1899, so I sorted the column by ascending and then removed the top 2 rows to remove them.
- The actual\_ETA column had some missing values, so I used the fill up option to fill the null values.
- I merged the ontime and delay column and turned it into a binary column which tells whether it was ontime or delayed, rather than telling it separately which causes the corresponding cell in the other column to be null. There were also some missing values so I used fill down to fill them.
- DestinationLocation and OriginLocation columns were split using a delimiter, in this case a comma, so that I could separate the Province from the whole location string and use it in analysis. Some of the values did not contain a province so I replaced them with null values and then filled down.
- Split the vehicle\_no column to only contain the first two characters of the string as they are repeated so they can be used in analysis.
- Split the origin and destination lat long columns so that latitude and longitude are in separate columns.
- The tripendtime column had some missing values, and as the actual\_ETA column is the same as tripendtime column so I replaced it with this column.
- There were some values listed as NA in the vehicleType and supplierNameCode column so I replaced them with Unknown in the columns.
- Dropped the columns that were not useful for analysis.
- Renamed some columns to more meaningful or appropriate names.

## Analysis of Charts and Story:

Query: How many bookings were delivered in each quarter of the given years and how many of those were delivered on time and how many were delayed?

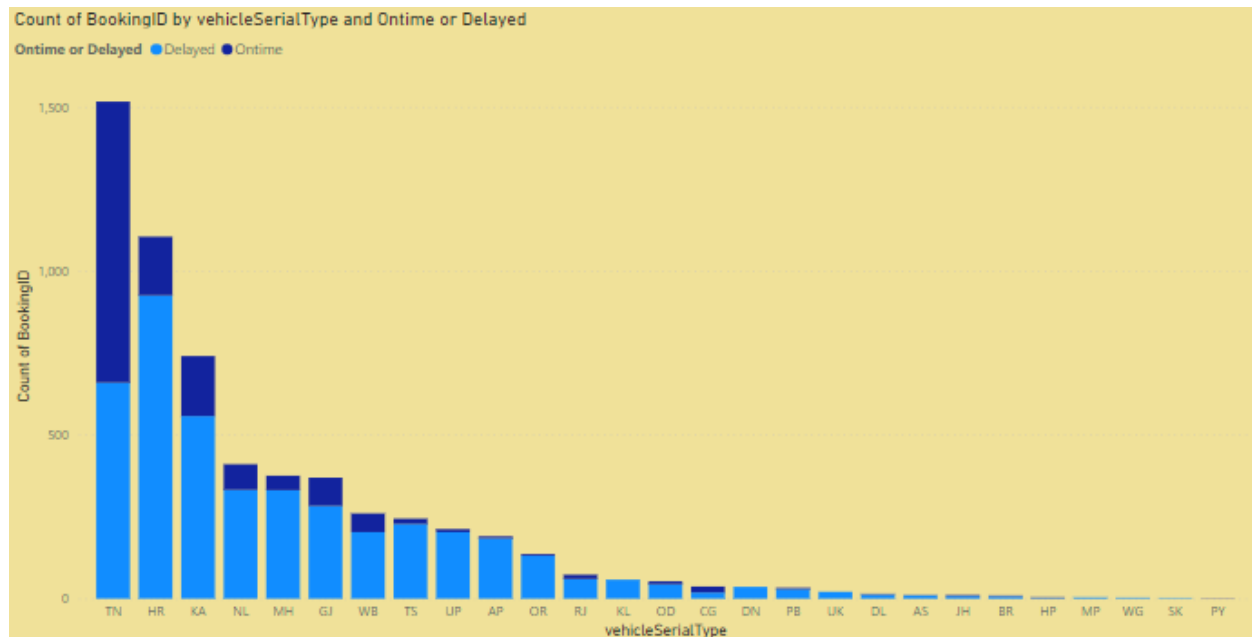


This is a stacked bar chart and it shows how many bookings were done in each quarter of the year (a quarter consists of 3 months) and splits them by whether they were delivered ontime or were delayed. As you can see from the graph, Quarter 3 (July, August, September) has the most number of bookings, while quarter 1 has the least. In all the quarters, there were more delayed deliveries than ontime deliveries. Quarter 3 was the only quarter that had somewhat the same amount of ontime deliveries as delayed deliveries, although delayed deliveries were higher. In the rest of the quarters the amount of delayed deliveries was significantly higher than ontime deliveries.



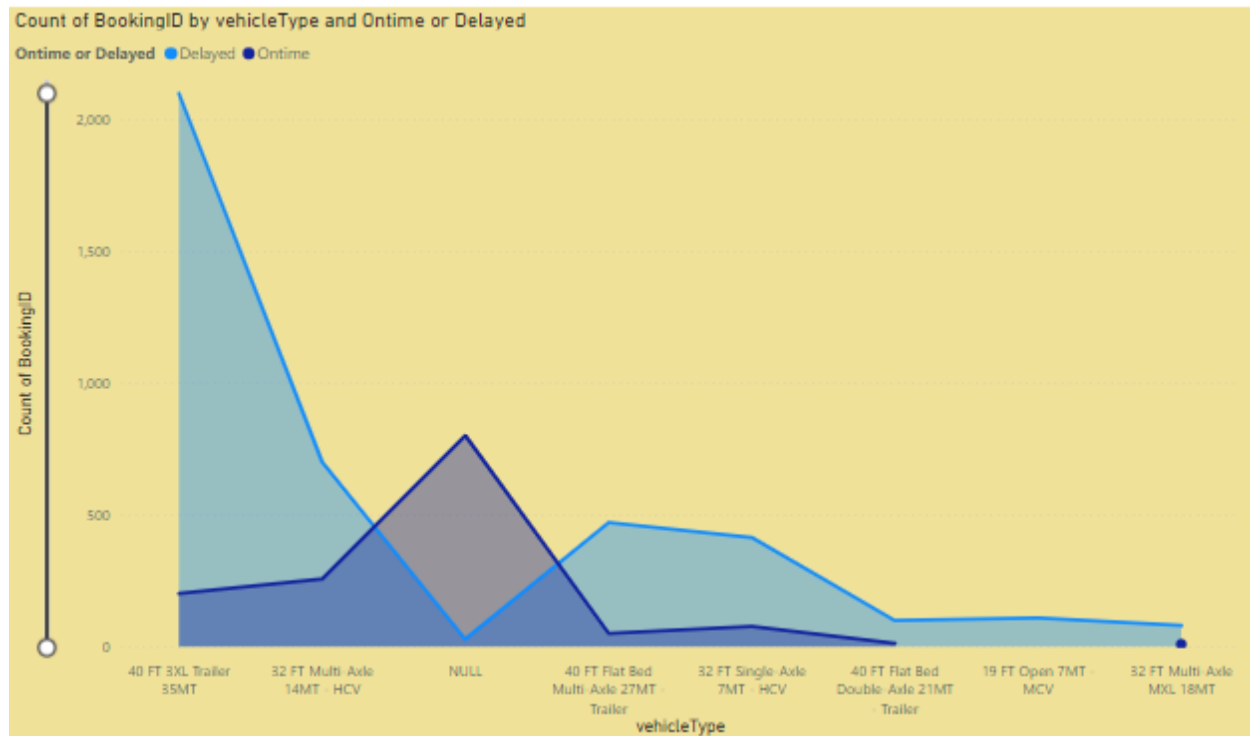
This is a decomposition tree and while I have shown above the number of bookings in a specific time period, this will show it in a hierarchical way such that first bookings will be divided by year, then quarters in that year, then months in that quarter, then days of the month. You can still see here that August is the month with the most bookings as it lies in Quarter 3, also known as the busiest month. We will also see this in the transportation distance charts below.

Query: How many bookings were delivered for each vehicle type with a specific serial number and how many of those were on time and how many were delayed?



This is a stacked column chart and it shows the number of bookings that were delivered with vehicles containing a specific serial type, split by whether they were delivered ontime or were delayed. As you can see from the graph, vehicles with the serial 'TN' had the most number of bookings as compared to other serial types. Also 'TN' was the only one which had more ontime deliveries than delayed deliveries, every other serial type had more delayed deliveries. This means that 'TN' is the most successful truck serial type and is the best choice if you are booking.

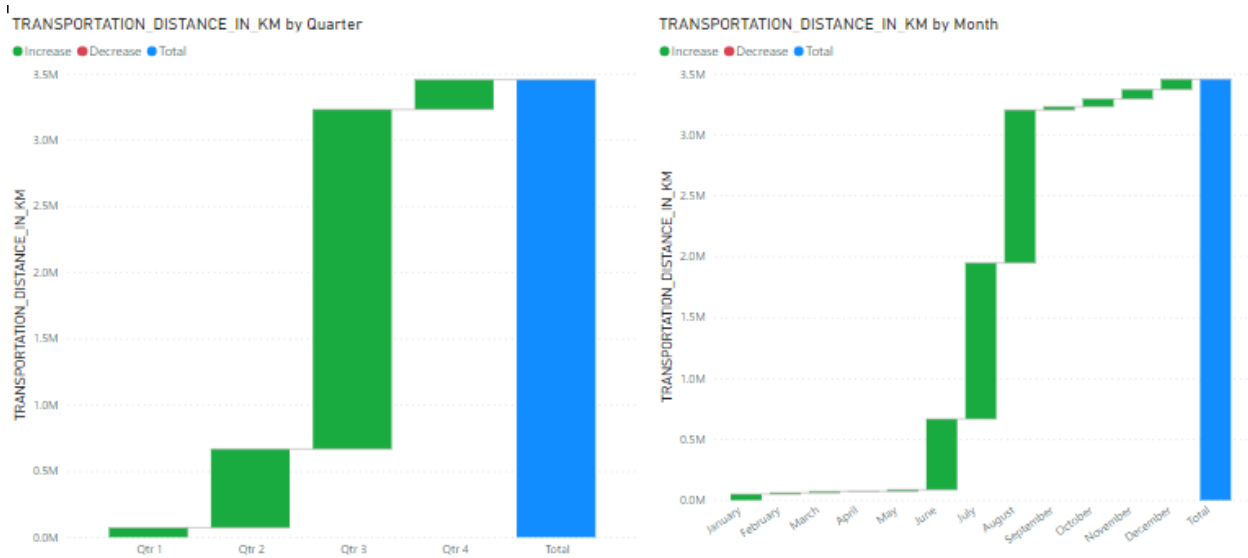
Query: How many deliveries were done through each specific vehicle/truck type and how many of these were on time compared to delayed?



This is an area chart and previously we looked at bookings on each serial type, but this time we will be looking at bookings on each individual vehicle type( I selected the top 7 for this graph ) split by whether they were delivered ontime or were delayed. We can see from the graph that the 40 FT 3XL Trailer was the one with the most bookings although it delayed the majority of them. At second, is the 32 FT Multi Axle truck, and this has about a third of the bookings of the first one, but this has more ontime deliveries, although the number of delayed deliveries is still high. There is also a truck of which we have no information of, and it has the highest amount of ontime deliveries and almost no delayed deliveries, if we had the name of this truck then it would be the best choice as it has the highest success rate. Almost all the vehicle types have more delayed than ontime deliveries.

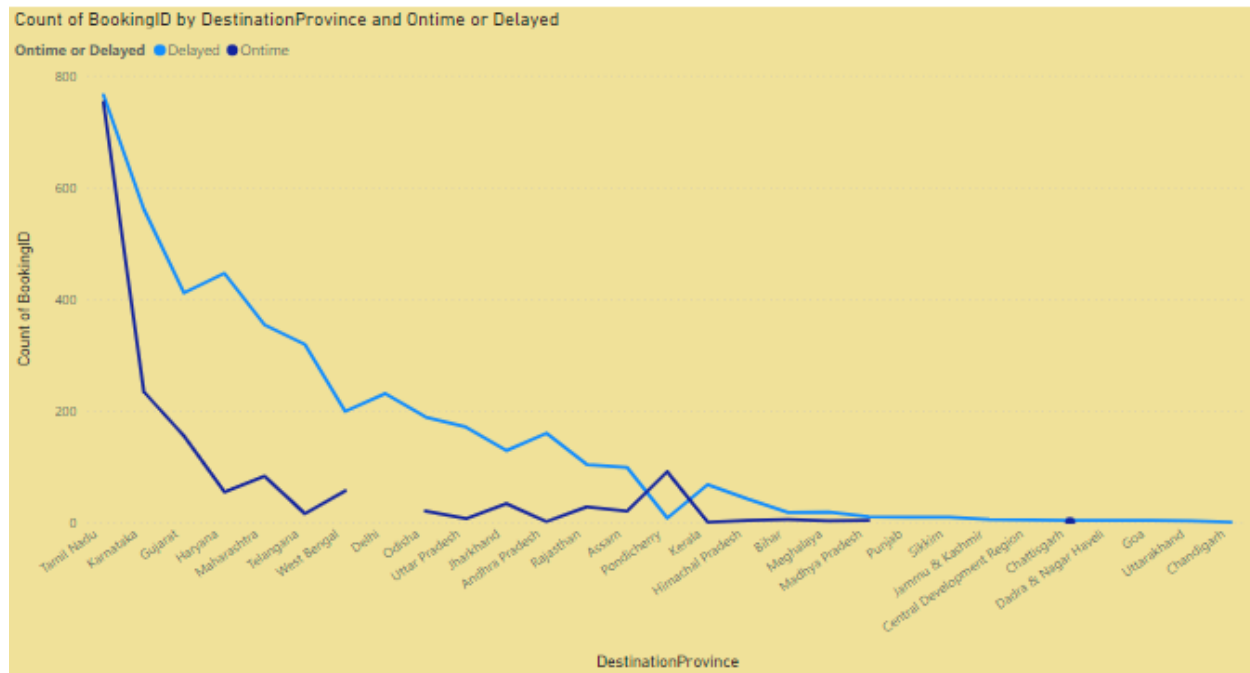


Query: What was the total transportation distance travelled by the trucks doing the deliveries in each separate quarter and each separate month of the given years?



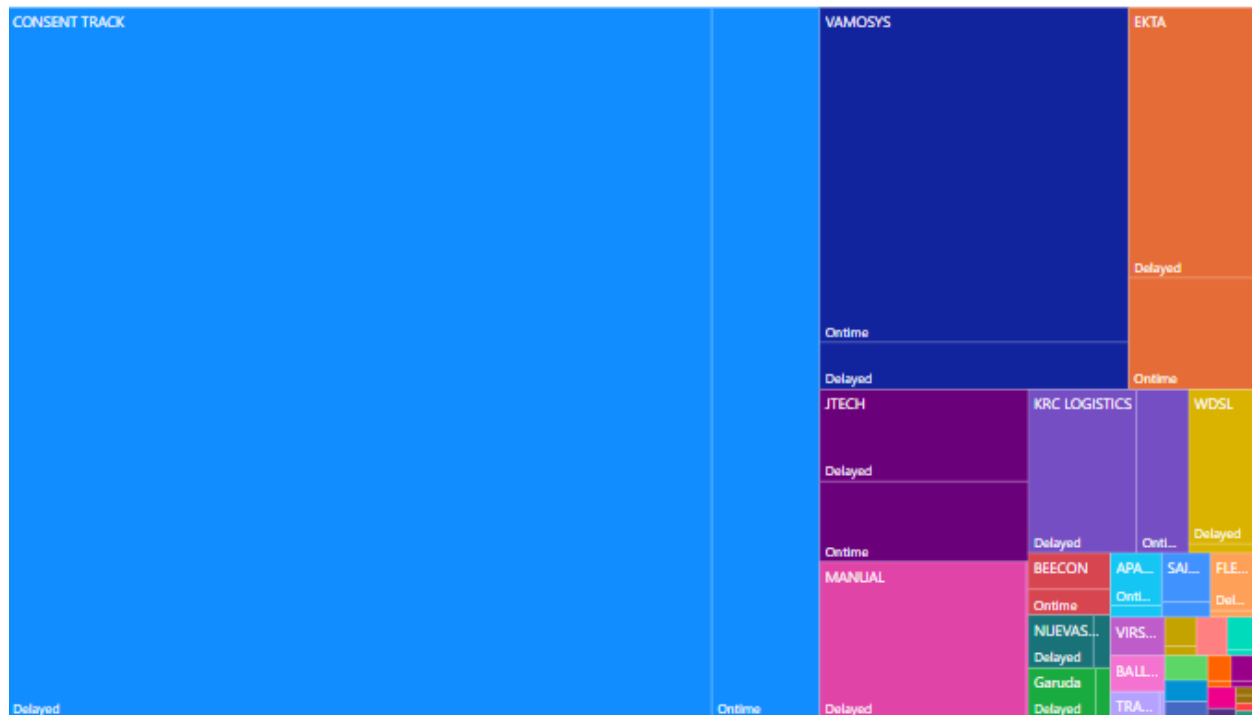
These are 2 waterfall charts and they are showing the transportation distance travelled by the vehicles, in the first chart by quarter and in the second by month. We have seen previously that most of the bookings occur in the 3<sup>rd</sup> quarter hence it makes sense that quarter 3 would have the most distance travelled. We can also see that the months July and August ( which lie in quarter 3 ) have the most distance travelled. Quarter 1 had the least bookings hence it has the least distance travelled. Quarter 4 had more bookings than quarter 2, but it has less distance travelled by the vehicle. This might mean that although the bookings were higher in quarter 4, the distance from origin to location might be less on average than bookings in quarter 2.

Query: Which destination province had the highest and lowest number of deliveries and how many of these deliveries were on time or delayed?



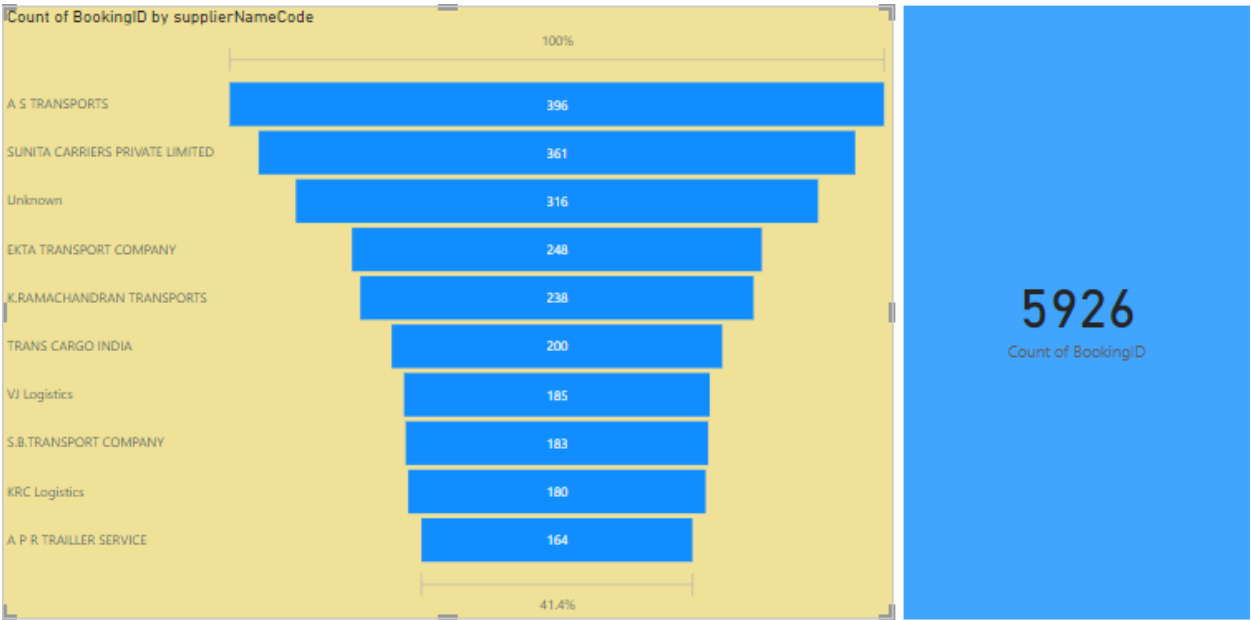
This is a line chart and it shows us the number of bookings and where they are to be delivered, also split by whether they were delivered ontime or were delayed. As we can see the highest number of bookings that have to be delivered are in Tamil Nadu. There are about 1600 bookings and about the same number were delivered ontime as they were delayed. Most of the provinces had majority of their deliveries be delayed, as we have noticed in the previous charts, with the exception of Pondicherry. The states of Delhi and Odisha had no ontime deliveries, all of them were delayed. This might be due to the fact that they are big provinces and might have a lot of traffic and delays, especially Delhi.

Query: Which GPS Provider was used the most by the trucks delivering and which GPS Provider had the highest successful delivery rate?

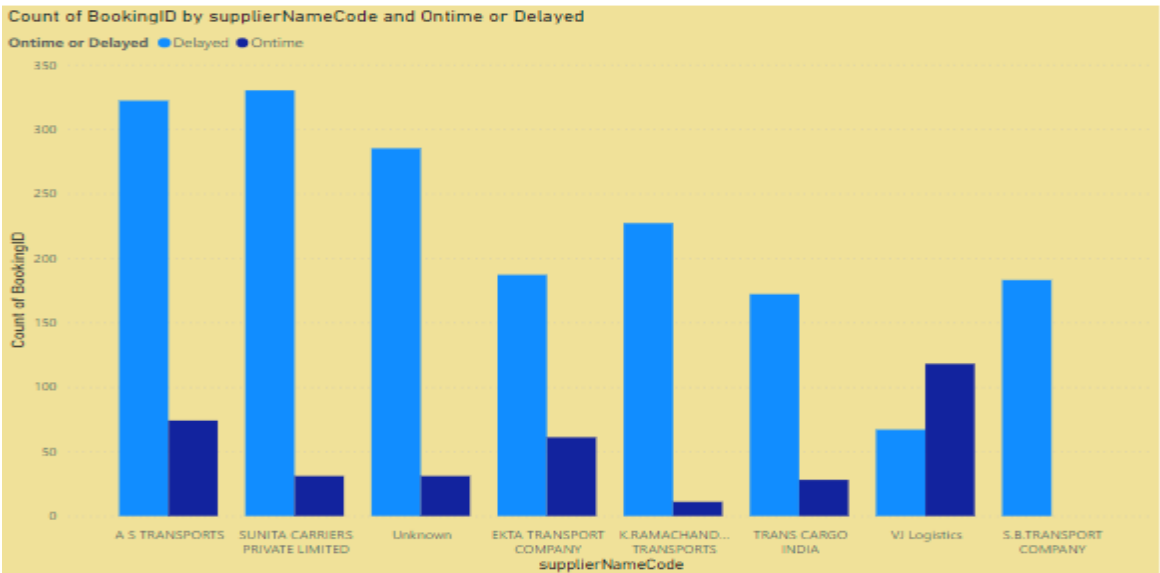


This is a tree map and it shows us the GPS provider used by the vehicle for the delivery and whether it was delivered ontime or not. We can see that more than half of the vehicles used 'CONSENT TRACK' for their GPS requirements. At second number is VAMOSYS and at third number is EKTA. About 80% of the bookings that used CONSENT TRACK had their deliveries delayed. Most of the GPS providers have more delayed deliveries than ontime deliveries. The exception is VAMOSYS, which has much more ontime deliveries than delayed deliveries, it has about a 85% success rate with deliveries. So, the best choice in regards to using a GPS Provider would be VAMOSYS. There were many GPS Providers which were used for less than 10 deliveries.

Query: How many trucks did each supplier provide for the bookings and which suppliers provided the highest amount of trucks and what was the success rate of deliveries done by trucks provided by each supplier?( Next 2 charts for this query )

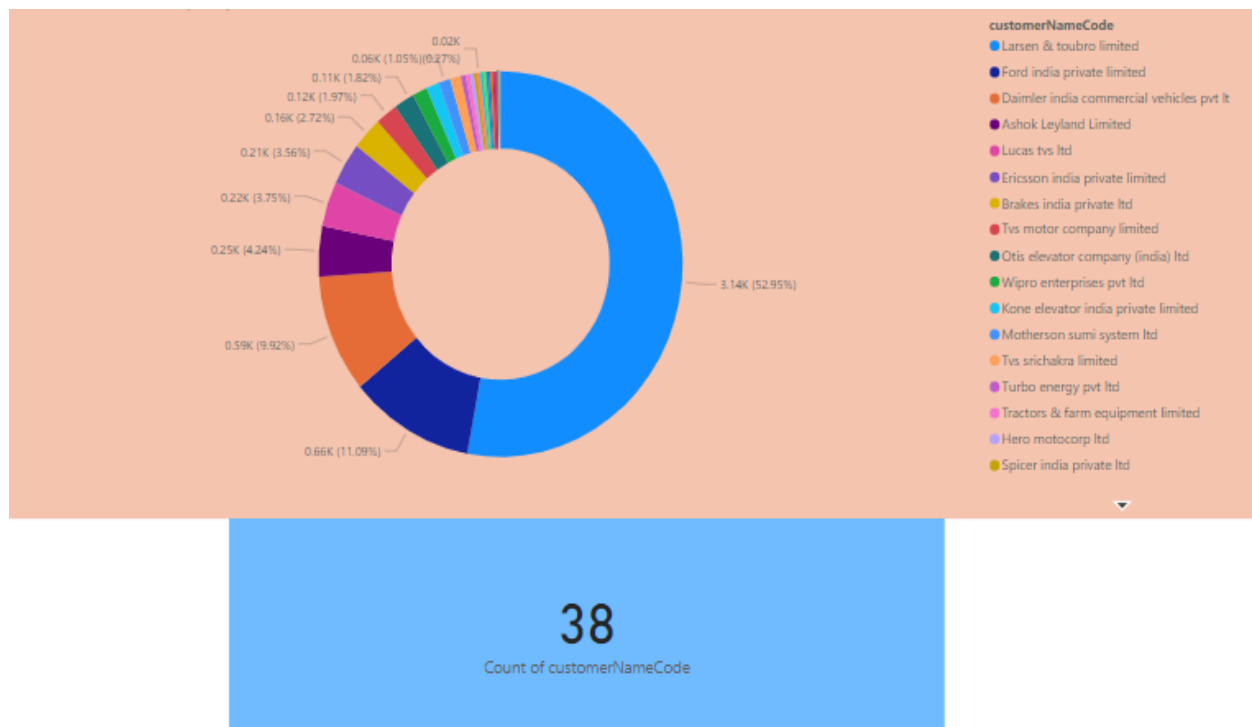


This is a funnel graph and it shows us the supplier that provided the vehicle for the delivery. I choose only the top 10 suppliers for this chart. The highest number of vehicles were provided by AS Transports, with 396 bookings going through them. Not far behind is SUNITA CARRIERS with 361 bookings. At 10<sup>th</sup> number in this graph, is APR TRAILER SERVICE, which has 164 bookings and this is around 41% of the highest supplier.



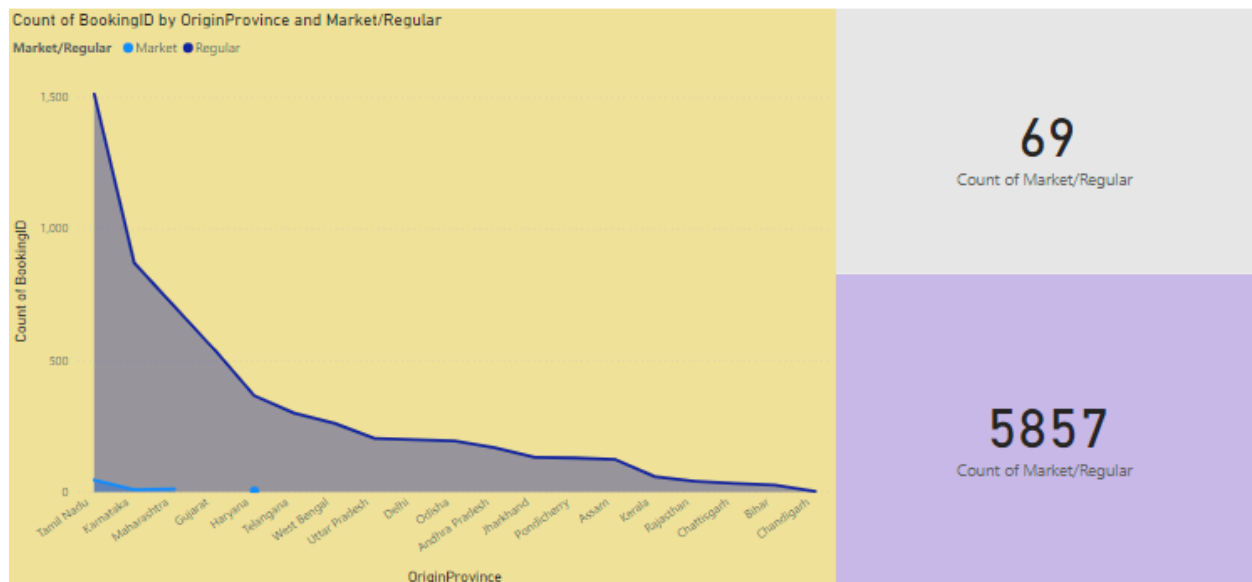
Following on from above, now we have those suppliers split into whether the delivery was ontime or not. As we can see most of the suppliers have a large proportion of delayed deliveries through their trucks, sometimes even 5 times as much. The only one which had a positive ontime to delayed ratio is VJ Logistics, they had more ontime deliveries. So, the best supplier would be VJ Logistics if we want the deliveries to be ontime most of the time.

Query: How many unique customers are there and how many bookings were done by each of the customers and which one had the largest share?



This is a donut chart and it shows us which customer had the most bookings. We can see from the chart that more than half the bookings are from a single customer alone, that is 'Larsen and Toubro Ltd'. There are 38 different customers but more than half of the bookings are just from one customer. And about 75% of all the bookings are from the top three customers, at second and third we have Ford and Daimler India respectively. The remaining 35 customers account for around 25% of the total bookings.

Query: Which origin province did the most bookings originate from and how many of these booking were market bookings as compared to regular bookings for each origin province?

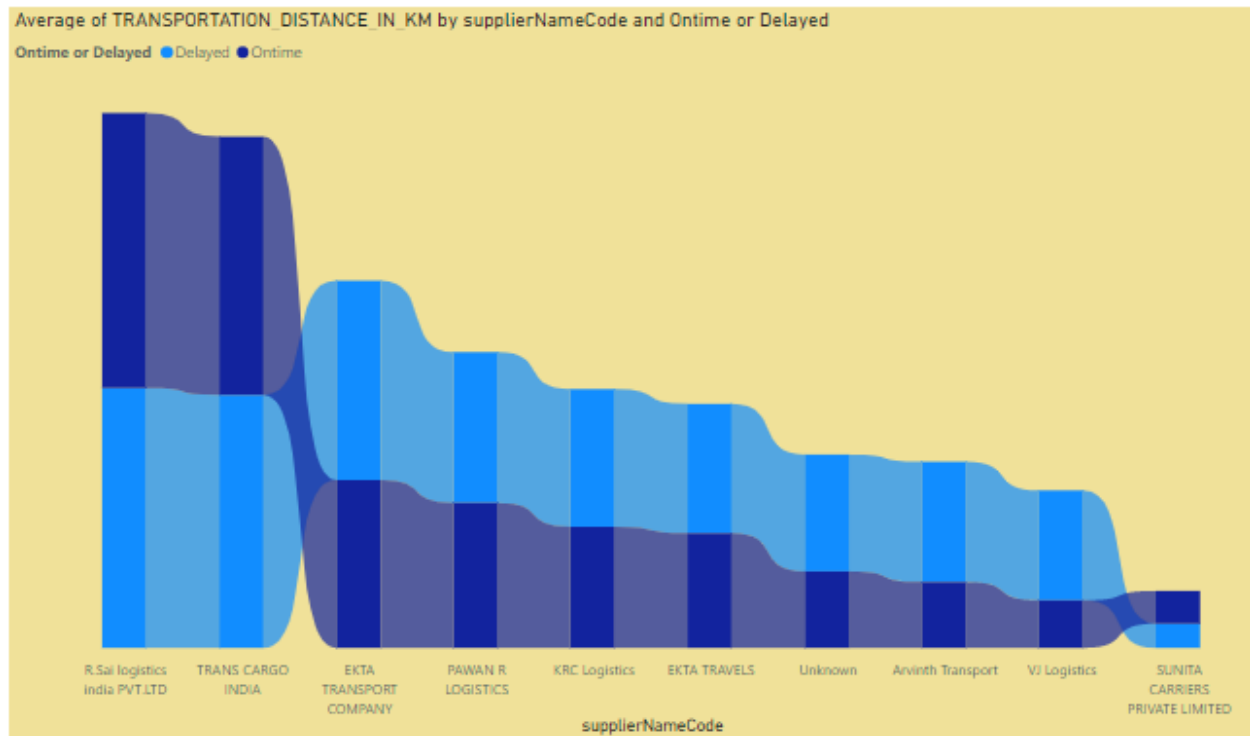


This is an area chart and it shows us the number of bookings by Origin Province split by whether they were market or regular bookings. In market the customer can have contact with the driver, while in regular the customer cannot have contact with the driver. As we can see from the chart, the overwhelming majority are regular bookings. Only 69 out of all the bookings are of market type and they were only done in 4 provinces, Tamil Nadu, Karnataka, Maharashtra and Haryana. Even these are very few when compared to regular. Rest of the provinces all solely regular booking type. We can further see the locations on the map below.



This is a map which shows the starting location of the trucks. I used the latitude and longitude coordinates given to build this map. As you can see most of the trucks start from the big areas and cities such as New Delhi, Mumbai and Chennai.

Query: What was the average distance travelled by trucks from each supplier, which ones had the highest average distance and what was the successful delivery ratio for each supplier?



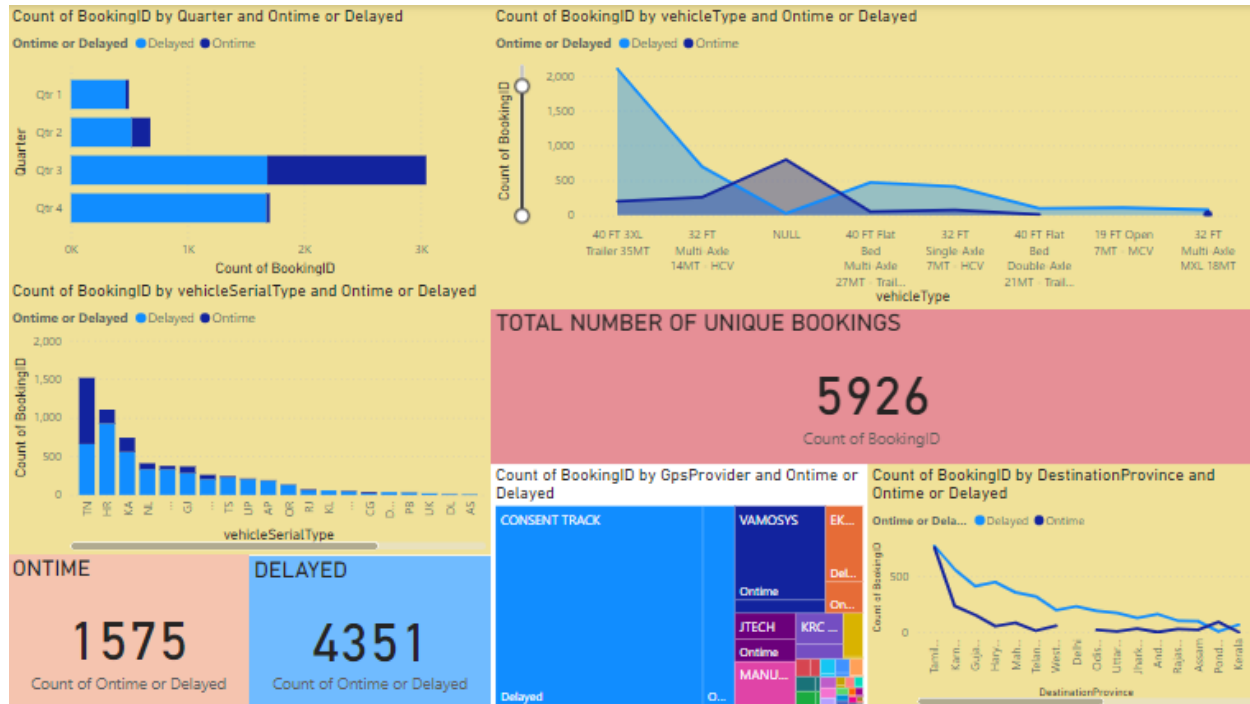
This is a ribbon chart and it shows us the average transportation distance travelled by the vehicles from different suppliers and whether the distance travelled was for an ontime delivery or a delayed delivery ( I have taken top 10 suppliers. We can see that the R.Sai Logistics and Trans Cargo had the highest average transportation distance but they are the only ones in the top that have a higher average transportation distance for ontime deliveries than delayed deliveries. The rest of the suppliers have a higher average transportation distance for delayed than ontime deliveries. This might mean that the shorter the distance from the origin to the destination, then it is likely that the delivery will be ontime rather than delayed, the reverse also holds true that the longer the distance, the higher the chance of being delayed.



## Dashboards:

### Ontime or Delayed Dashboard:

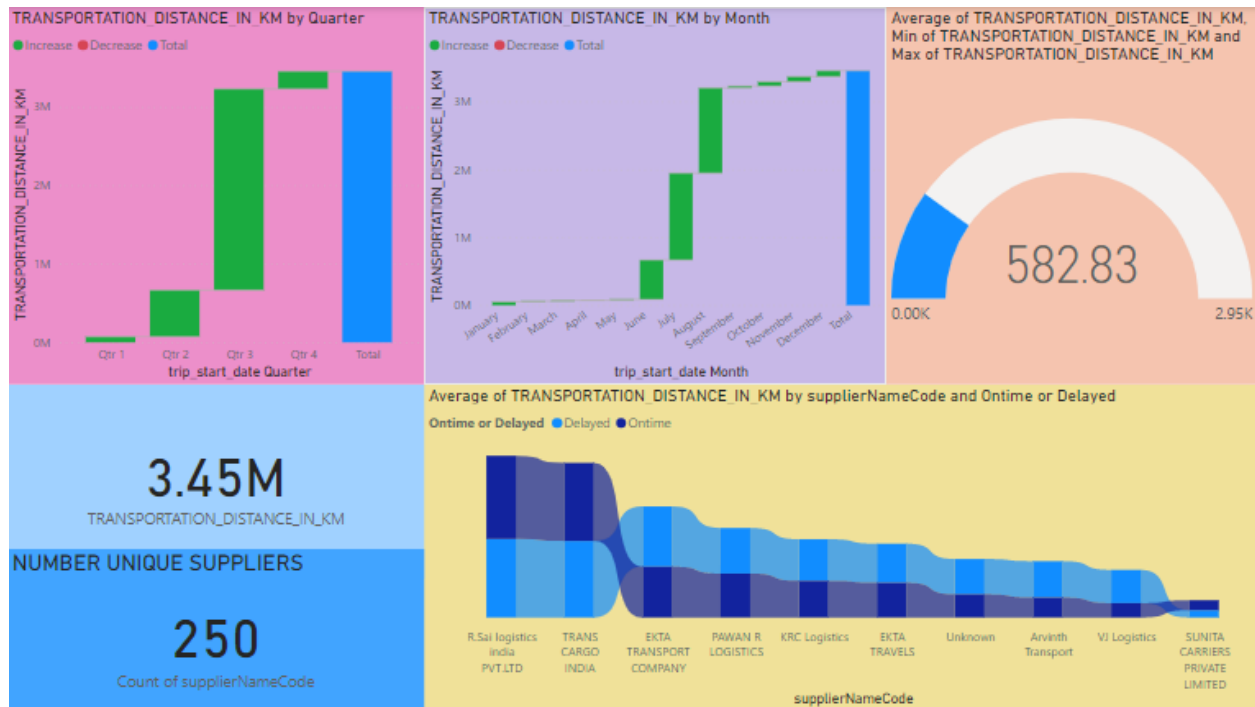
Query: Ontime or Delayed across different dimensions



This dashboard shows whether the bookings were delivered ontime or they were delayed and measures it across multiple different dimensions.

## Transportation Distance Dashboard:

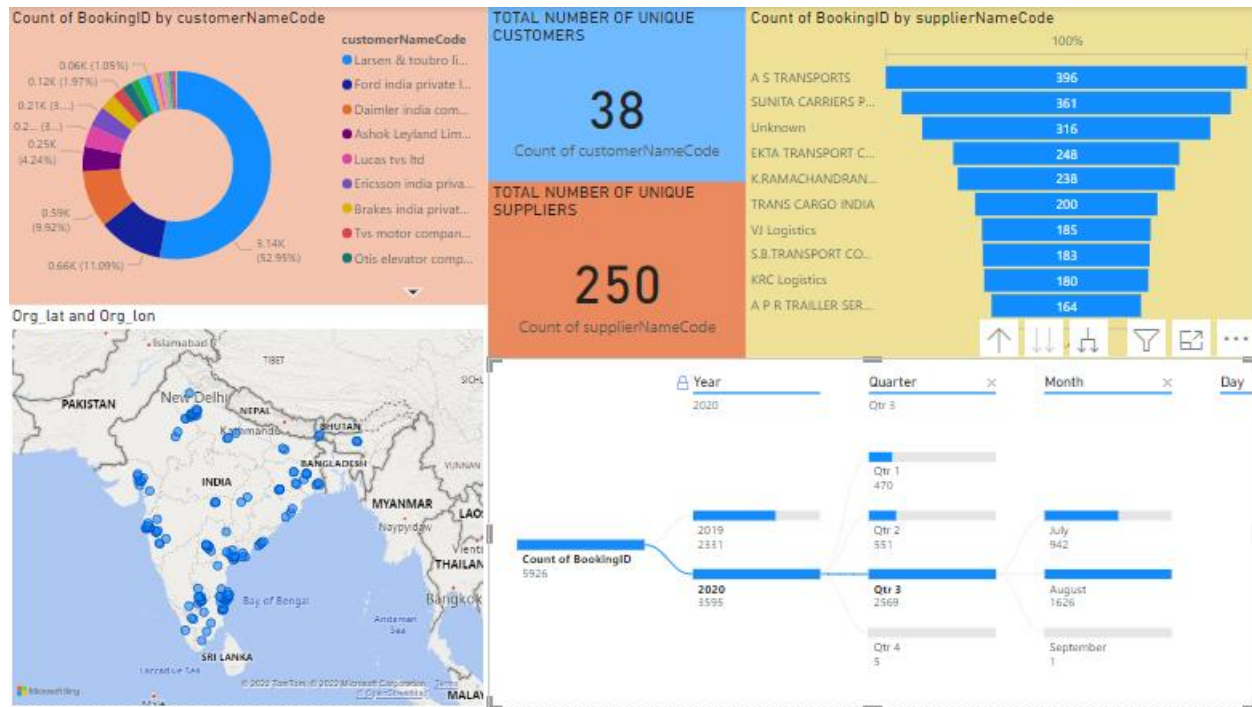
Query: Transportation Distance across different dimensions



This dashboard shows the transportation distance travelled by the trucks while making the deliveries and measures it across multiple dimensions.

## Bookings Dashboard:

Query: Number of bookings or deliveries across different dimensions



This dashboard shows the number of bookings or deliveries and measures it across multiple dimensions. You can see that in the dashboard above we have done it by customers, suppliers, location, time etc.