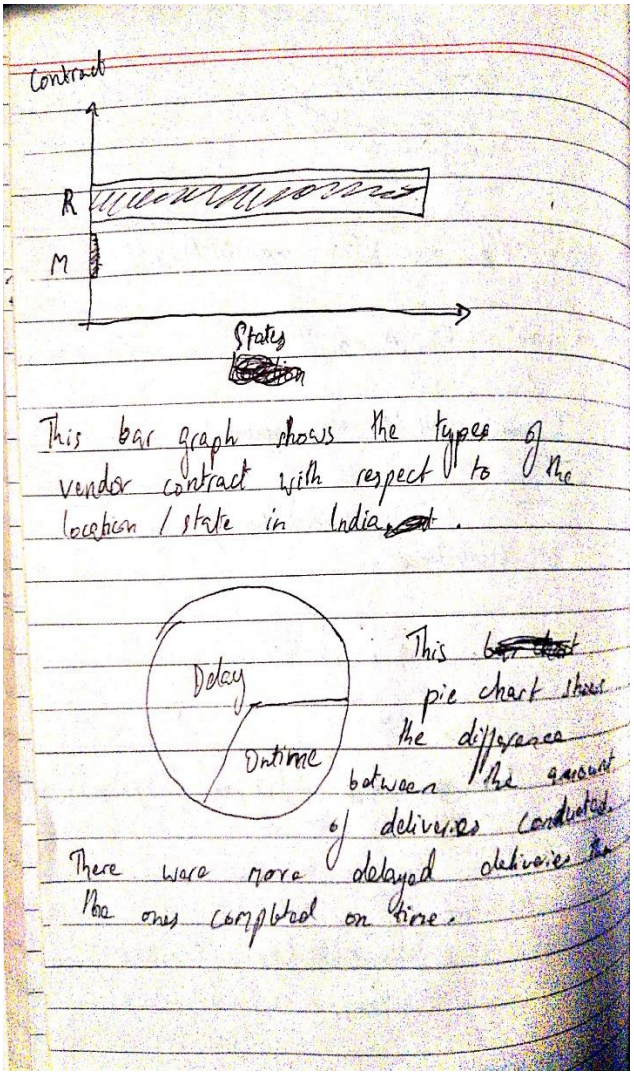
BI Project

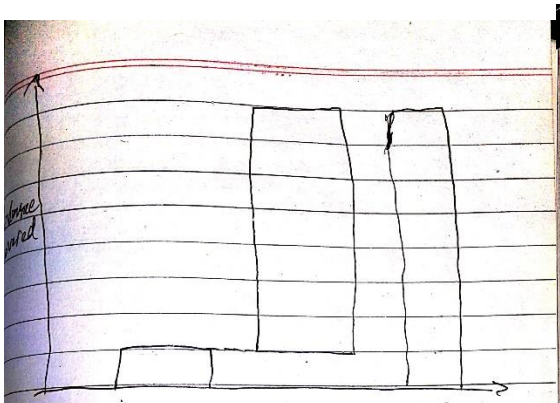
Obaidullah – 18615

Dataset used: Delivery truck trip data

Tools used: Most of the wrangling and simple visualizations were done on Python. The rest of visualizations were done on Power BI for deeper analysis.

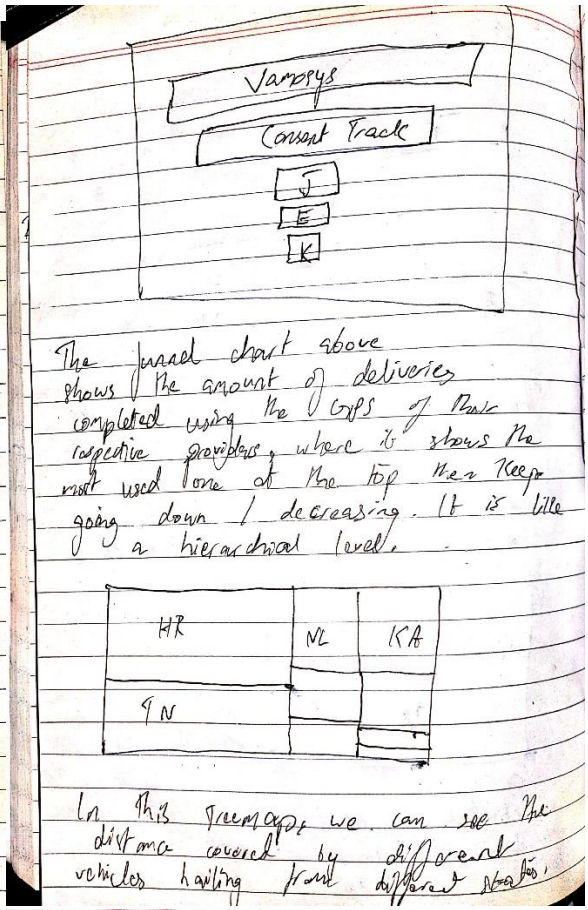
Paper Charts





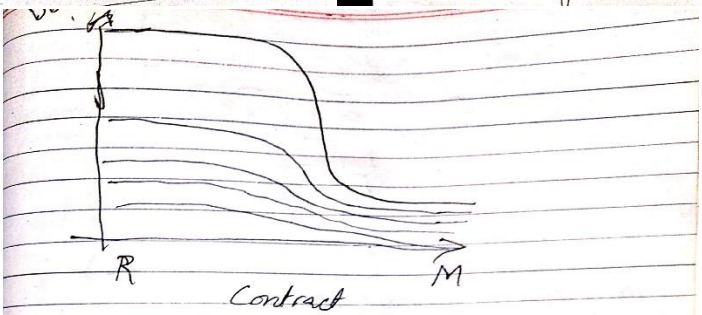
The waterfall chart above shows the distance travelled by all the trucks within the two years. 2019 had covered significantly lesser distance than 2020.

There was huge increase of transportation rate very next year.

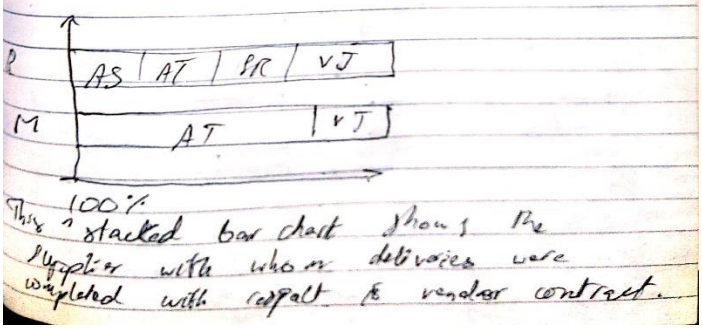


The treemap chart above shows the amount of deliveries completed using the trips of their respective providers, where it shows the most used one at the top then keeps going down / decreasing. It is like a hierarchical level.

In this treemaps, we can see the distance covered by different vehicles hailing from different states.



The ribbon chart shows the disparity in amount of deliveries with respect to the type of vendor contract, where deliveries made through vendor contracts are predominant.



This 100% stacked bar chart shows the supplier with whom deliveries were completed with respect to vendor contract.

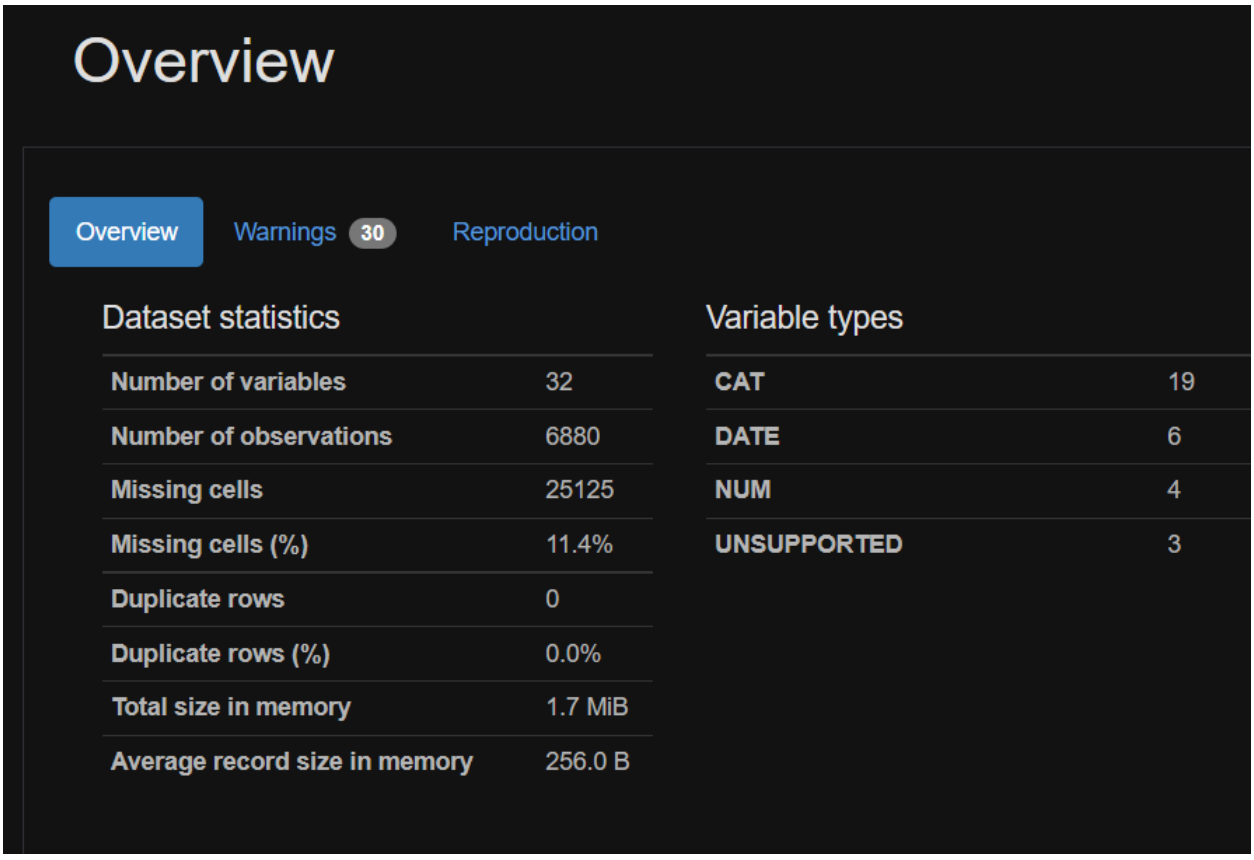
Data Transformation:

I used Python for effective wrangling for this dataset.

data=pd.read_excel(r'C:\Users\lenovo\Downloads\BI_TruckDelivery\Delivery truck trip data.xlsx')

Numpy, pandas, matplotlib (plt.style.use('ggplot')), geopy, and pandas_profiling packages were used. Firstly, to get a glimpse on the overall situation of our dataset, we use profiling. However, it wasn't that straightforward. Not because it was hard to implement, rather importing pandas_profiling on the later pandas version was creating problems and wouldn't let us get past, as it kept throwing an *Import Error*. There are two solutions, downgrading your pandas to 1.2, or going into your *conda* directory, and finding out the *Boolean.py* file, and changing ABCIndexClass to ABCIndex. I went with the latter approach which fixed the issue.

pandas_profiling.ProfileReport(data)



This provided thorough profiling of each column, whether they had missing values, high cardinality, or unsupported data.

We then spot an anomaly in time-series data, where two rows had the year 1899 on them after sorting them in ascending order. This is a problem since our actual data starts at 2019.

data.sort_values('Planned_ETA').head(5)

So we dropped both of these indexes.

BookingID and BookingID date did not have much to do with our analysis, so those columns were deleted.

Market/Regular column was renamed to Vendor_Contract.

Ontime and delay columns had only singular values consisting of G and R respectively. Therefore, I merged them and made G = 0, and R = 1.

Vehicle_no can be identified which state they were from by their first 2 characters, therefore it was converted into string and vehicle_state column was formed.

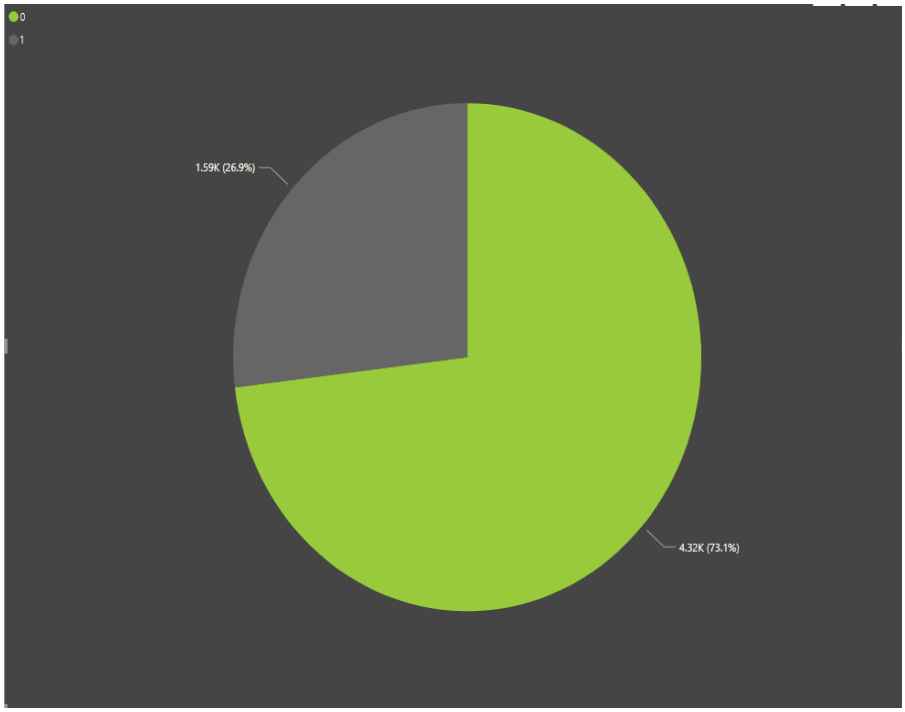
Driver_MobileNo; there were too many missing values, so my strategy to amend this and form it into something useful was to check whether the driver's mobile number existed by applying a condition that if the value exists and is greater than 0, we assign it as 1. All the NULL values were assigned as 0.

Origin_Location and Destination_Location had a lot of elongated values with respect to their states, which appeared to be redundant, so they were trimmed and made only to show their states, therefore another two columns were made called Origin_States and Dest_States.

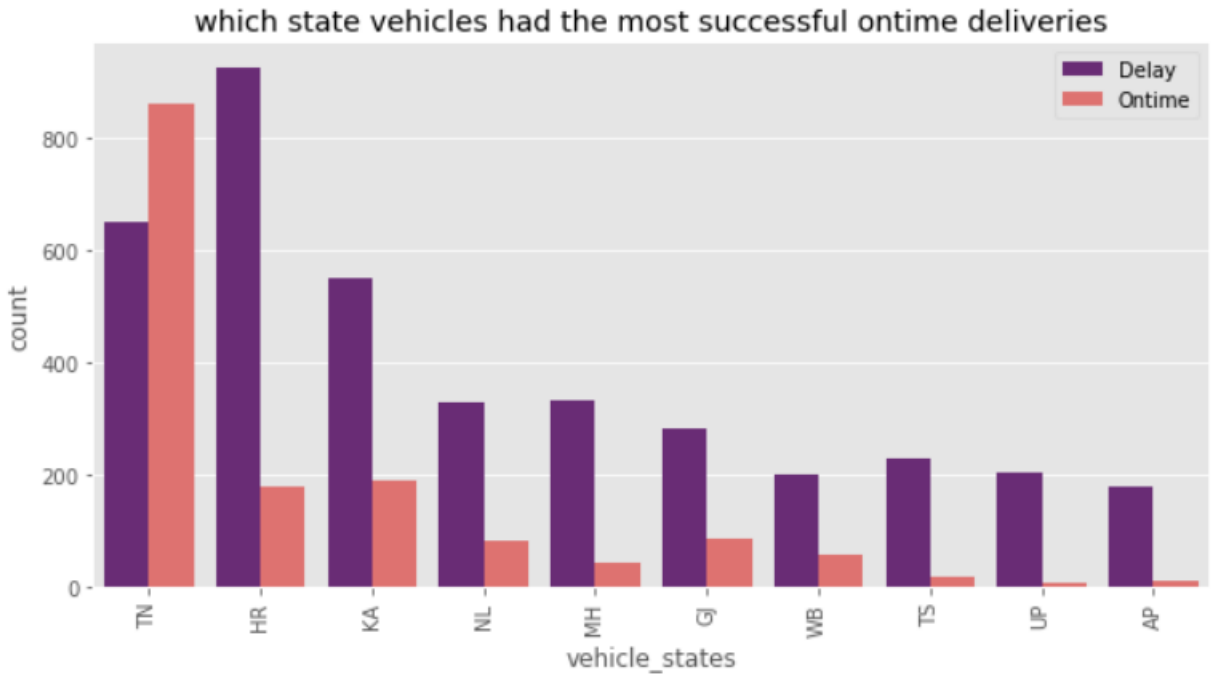
After completing the visualizations in Python, this command was used to generate an output excel file for further wrangling and deeper analysis using a different visualization tool. I used Power BI to perform complex visualizations to answer business problems.

```
writer = pd.ExcelWriter('Delivery truck trip data.xlsx', engine='xlsxwriter')

data.to_excel(writer, sheet_name='Python_Wrangled')

writer.save()
```

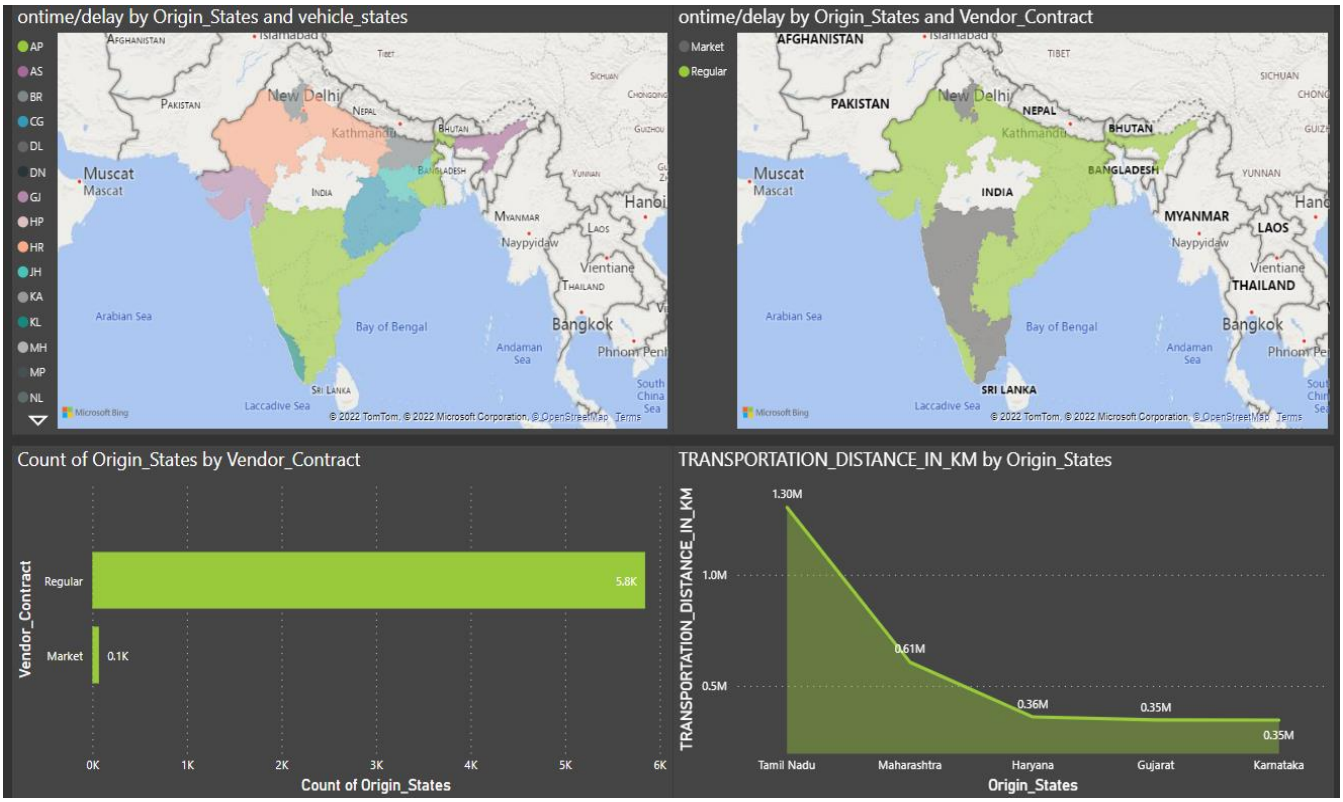
Visualizations & Analysis

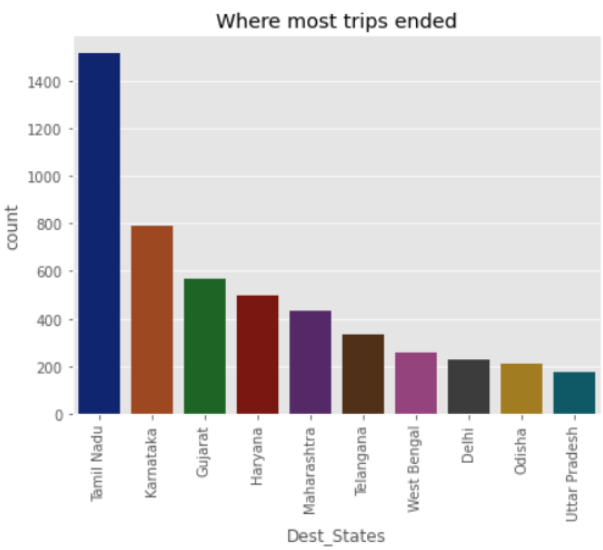


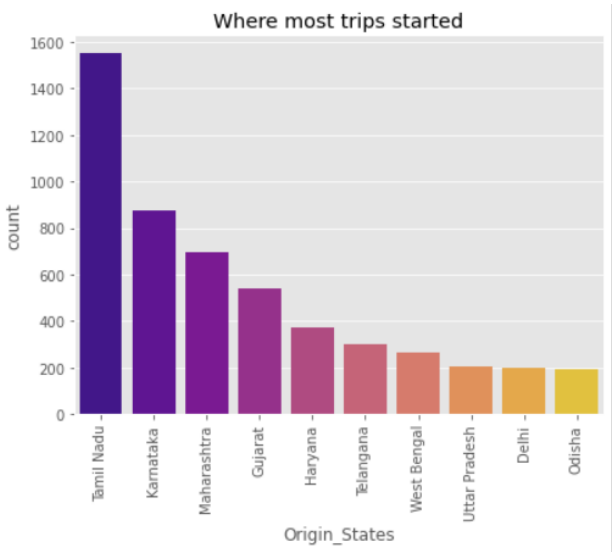
In the pie chart above, we can see that most of the deliveries were delayed. While showing a pie chart at first may not be a plausible strategy for a business analytics presentation, it can be used to establish a fact or boundaries that may exist and play a significant role in our later analysis. 73% of the deliveries were delayed, while only 27% deliveries were made on time.



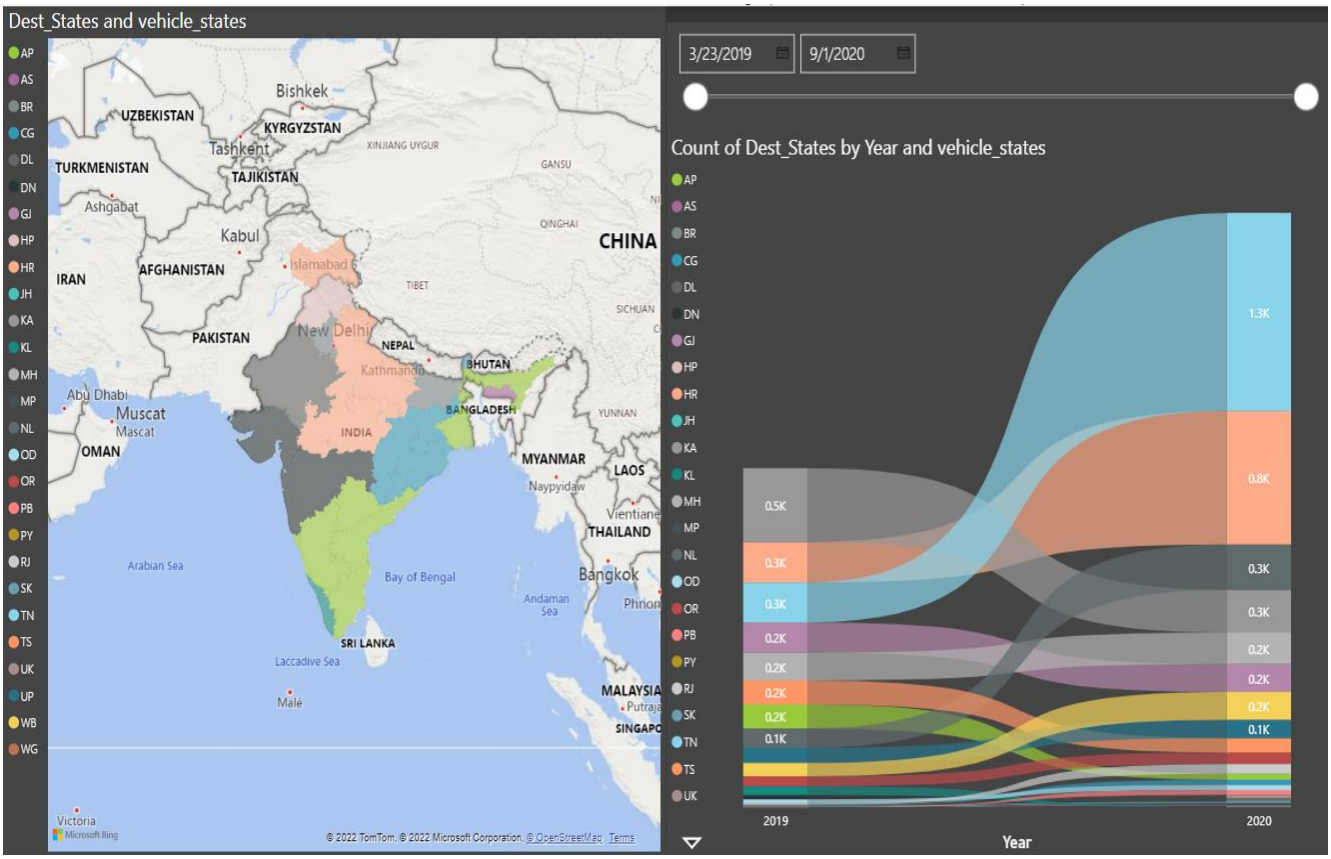
- We can see that the trucks hailing from Tamil Nadu had the most successful on time deliveries, and overall also the most successful deliveries, compared to all the other state trucks which mostly resulted in delayed deliveries. Trucks hailing from Haryana had the most delayed number of deliveries.



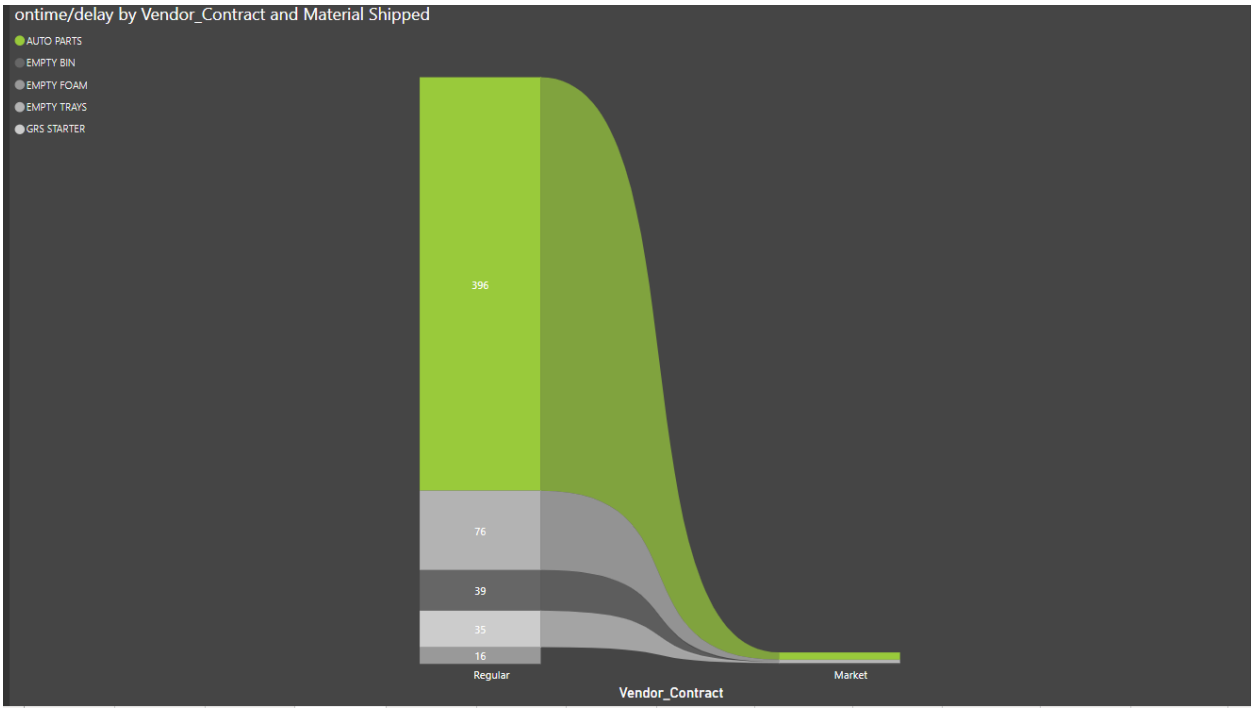


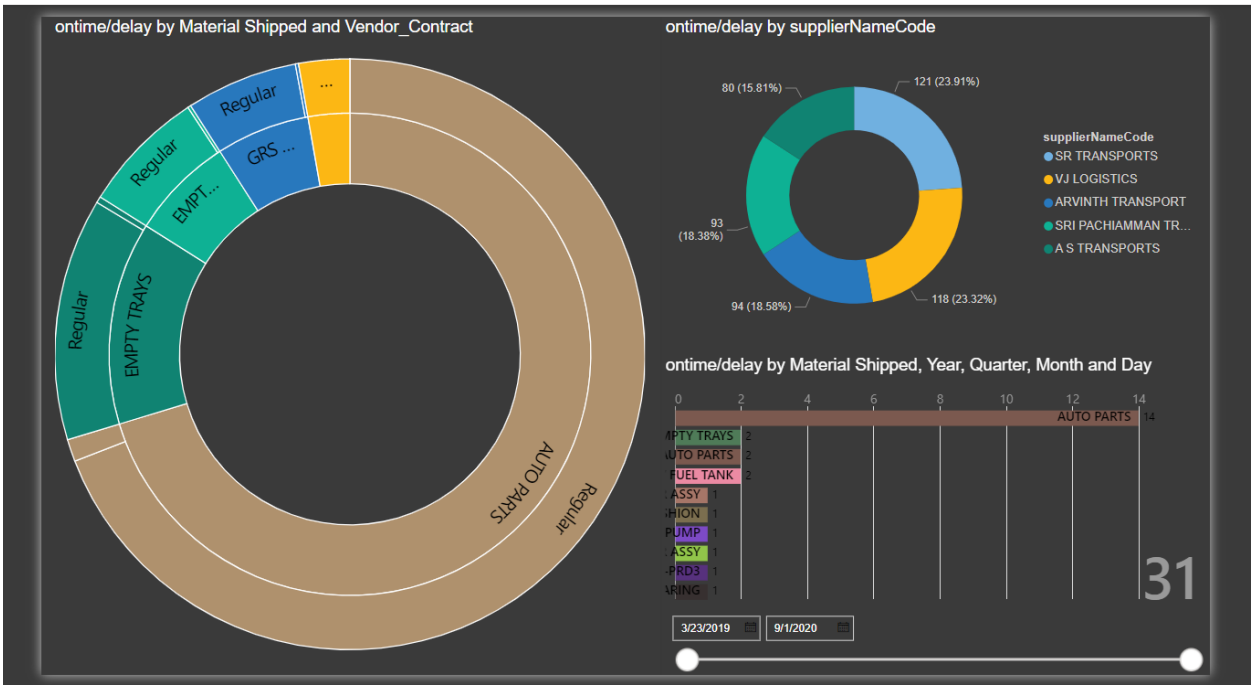


- Most trips started and ended in Tamil Nadu itself, whose truck models also had the most successful number of on-time deliveries. This could be one of the major factors in its successful deliveries, since the same truck models were utilized in the same states, therefore the chances of successful on-time deliveries rises exponentially.
- Trips started from Tamil Nadu also has the greatest distance covered/farthest amount of kilometers travelled.
- We can see that most of the states of India where the trips started, Regular contracts were the most common, which means that it was common for truck companies to have vendors with whom they have contracts overall.
- A vendor contract is essential for a logistics or delivery business, since it establishes a formality where there would be proper compensation with respect to the outcome. Therefore, it is only rational that it completely overshadows the vendors without contract, since it is more risky for the business.



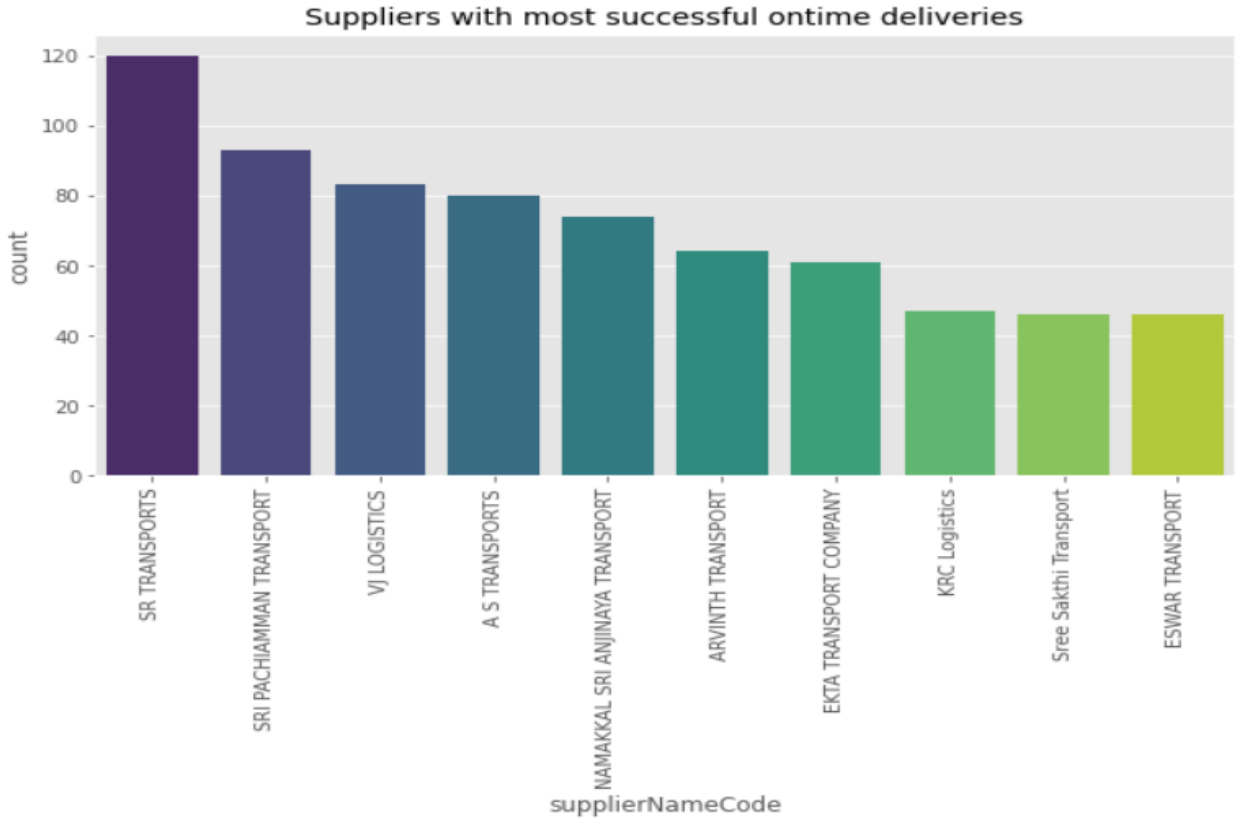
- Trips that ended by different truck models originating from their states respectively can be seen in the ribbon chart above, and how it varies through passing time. We can see that trucks hailing from Tamil Nadu was not as popular as trucks hailing from Karnataka and Haryana, however as time passed, the use of Tamil Nadu trucks grew exponentially, not just surpassing the two above but completely overshadowing them. Haryana trucks also overtakes Karnataka's, only following Tamil Nadu trucks now. Nagaland trucks have made a huge improvement, from being so behind in 2019, to being the top third state truck in 2020.
- We can see in the ribbon chart below the disparity which has further been accentuated here between the vendor contracts, with respect to the deliveries and the materials shipped by each. It is obvious that Auto Parts were the most prominent in this scenario for both the vendor contracts.

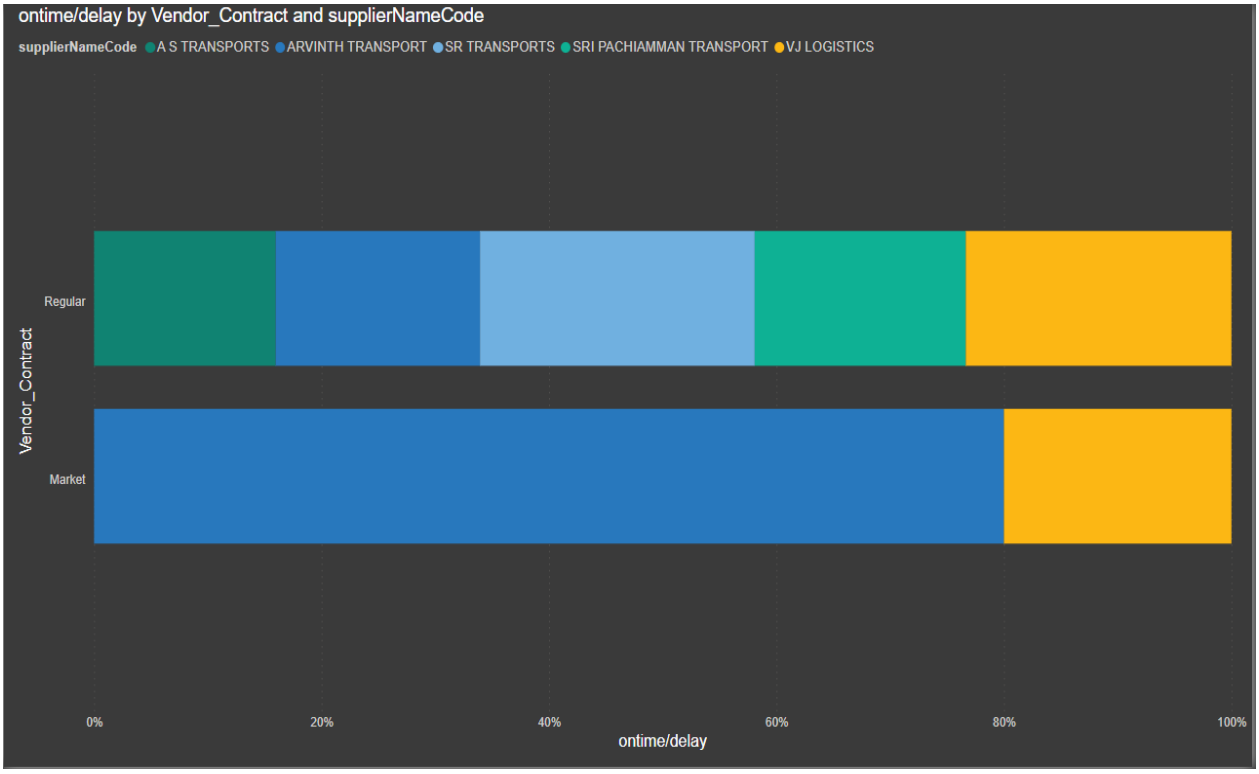




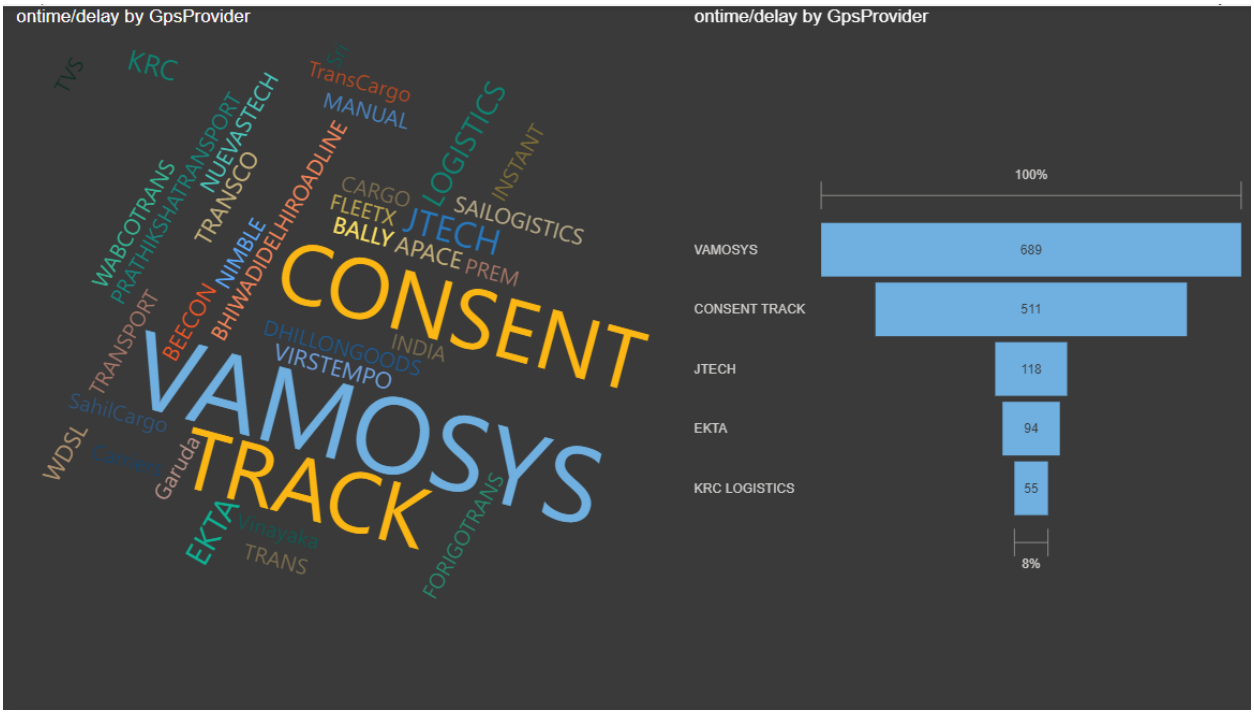
- In the sunburst chart, we can see different items that were shipped along with their deliveries more clearly, along with whether they had a vendor contract or not. Auto Parts covers over 70% of the materials shipped, which shows that the truck deliveries or the logistics company that uses these trucks, is quite likely a business where they repair or conduct maintenance for vehicles of other customers. This also becomes prominent by the fact that other items that follow it are also for vehicles, except empty bins which is general use.

- We can see that most of the deliveries are completed by SR Transports, followed by VJ Logistics who are just three deliveries short.
- In the bar chart below, we can see the number of successful deliveries on-time with respect to each supplier. We can see that SR Transport has been highly successful as they completed all but one deliveries on-time, which is a huge success (approximately 99% deliveries completed on time).
- Sri Pachiamman Transport are also very successful in this regard since they have completed 100% of their deliveries on-time, despite not having the highest number of deliveries, it's still among the most reputable.
- VJ Logistics may have the second-most number of deliveries finished, their performance was rather poor, since they only completed approximately 70% of their deliveries on-time.

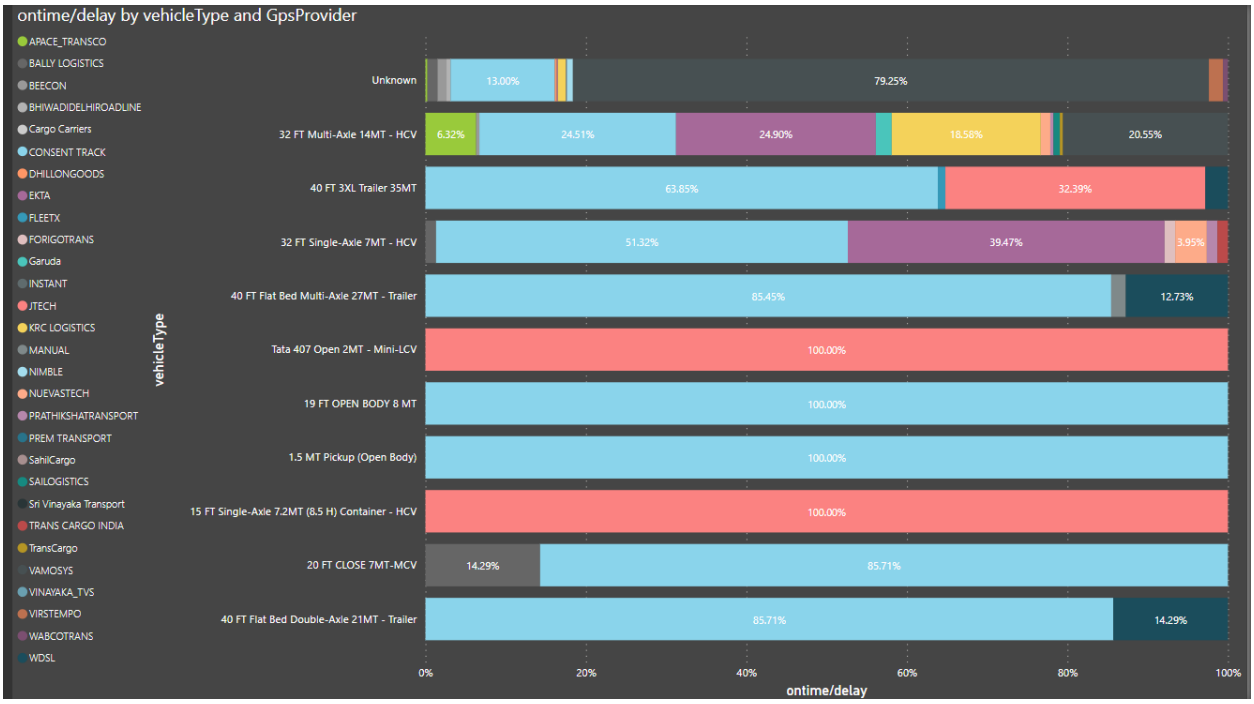


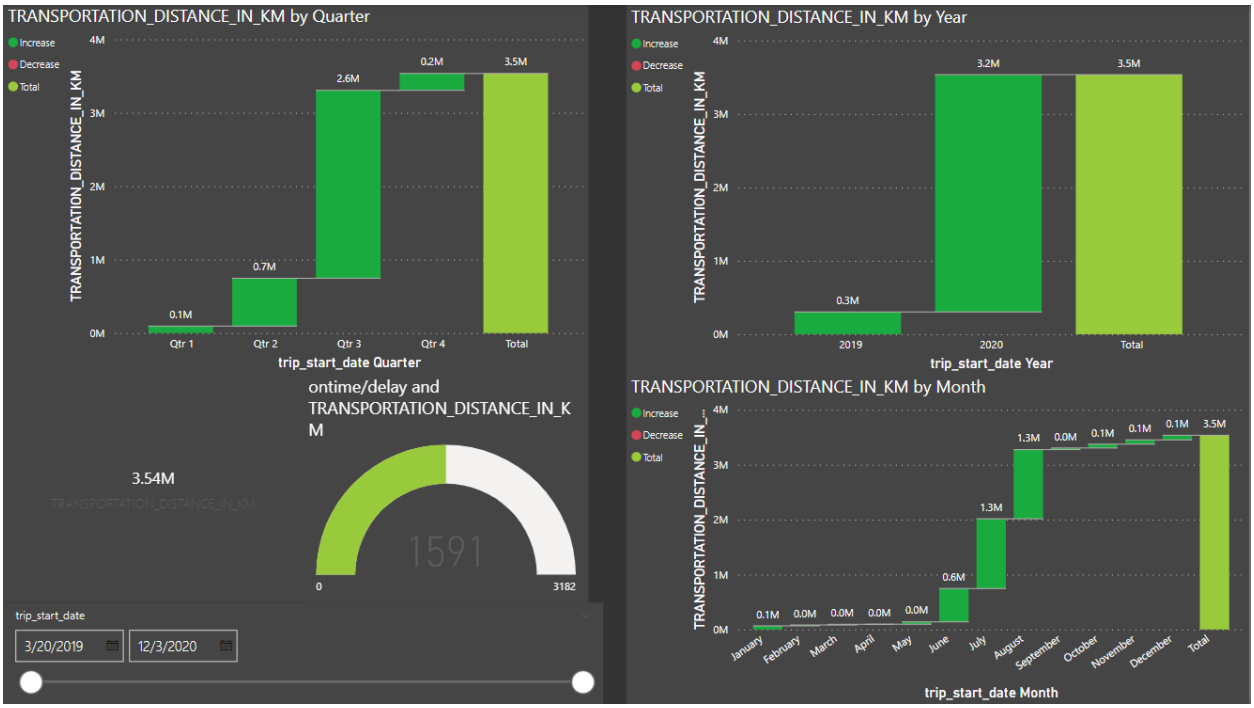


- The 100% stacked bar chart above has the answer as to why. The yellow highlighted one belongs to VJ Logistics. This shows that they have vendors with whom they do not have a vendor contract with, which means that in case of untimely delivery, vendors may not face any specific legal ramifications, therefore the probability of tardiness and accidental damage is higher than the ones with whom there exists a vendor contract.



- There must be a preferred GPS provider that must be prominent in logistics for truck deliveries, in order to effectively complete the deliveries while reaching the correct location. We can see that Vamosys is the most prominent GPS provider, therefore most trusted amongst the logistics and trucking companies.
- Consent Track is a close competitor in this field. This could mean that they both might offer various credits and discounts in relation to bulk-buying, sponsorship etc. to their customers in order to remain competitive. The rest of the companies do not have that much prominence in this scenario and fall short.
- In the 100% stacked bar chart below, we can spot an anomaly that put Vamosys in the most popular GPS provider, where most of it is supplied to “unknown”, therefore makes our analysis redundant. Consent Track seems to be dominating in otherwise all known truck models for performing deliveries. So we can safely assume that undoubtedly, Consent Track is the most reliable GPS provider.

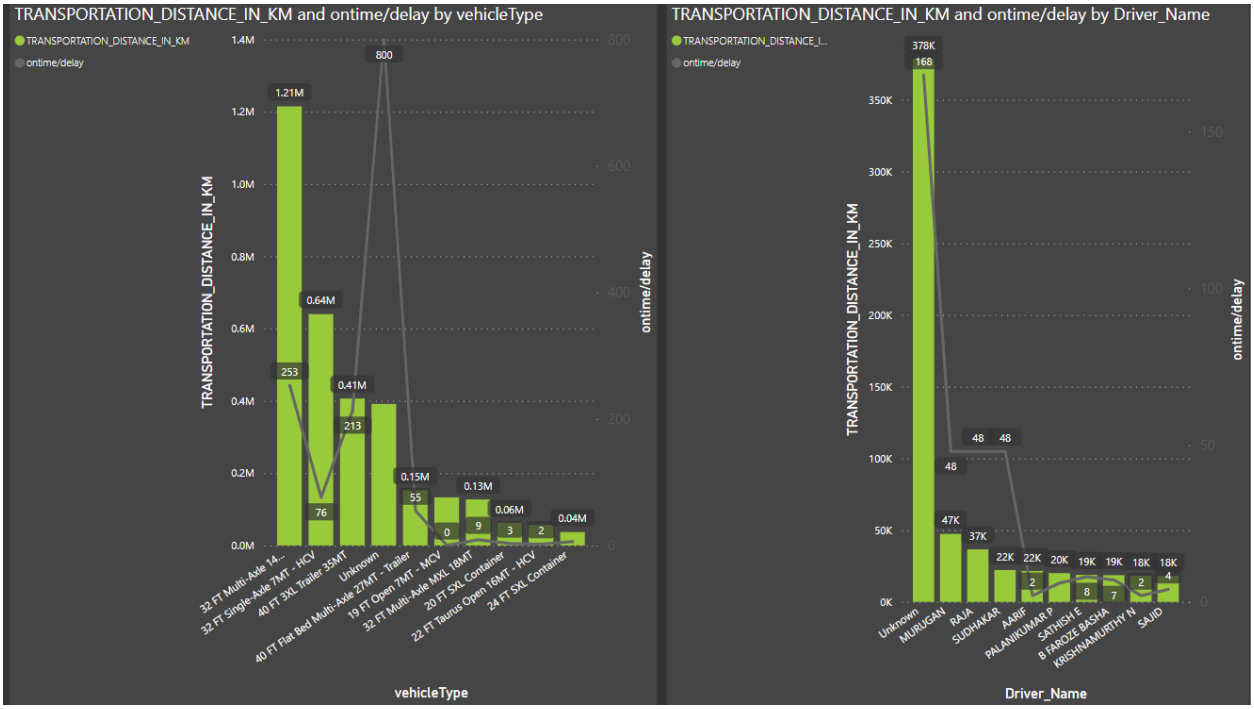




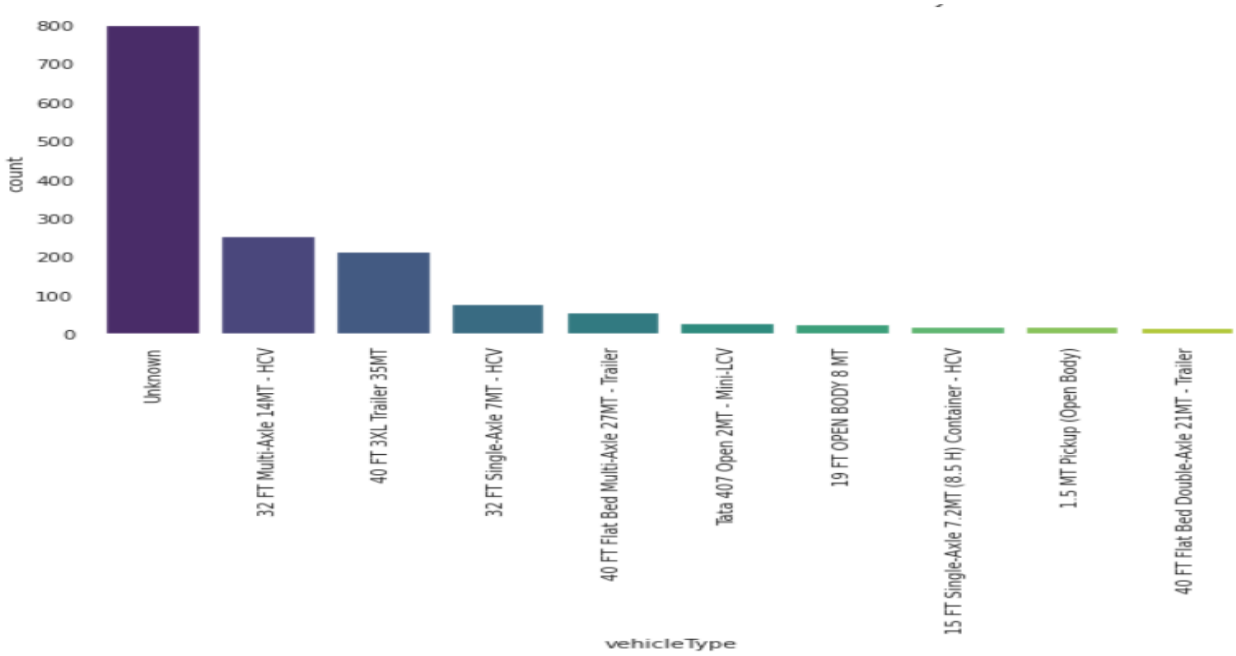
- In the waterfall charts, we can see the total transportation distance covered in each year, month, and quarter. Similar to the graphs above, I have attached a slicer that can be used to alter time and check the variations that may occur in the distances covered.

- Distance covered in 2019 was miniscule to that of distance covered the very next year, which has a sharp increase.
- The greatest number of transportation distances are covered mostly in the summer season, and the third quarter is indeed the busiest.
- We can see that the total distance covered is 3.54 million kilometers, out of that, approximately half of that distance was used to complete the deliveries. Which could lead to a serious business problem since the consumption of diesel would be a waste of resources, and therefore the truck or logistics business will incur more costs, probably leading to a loss.

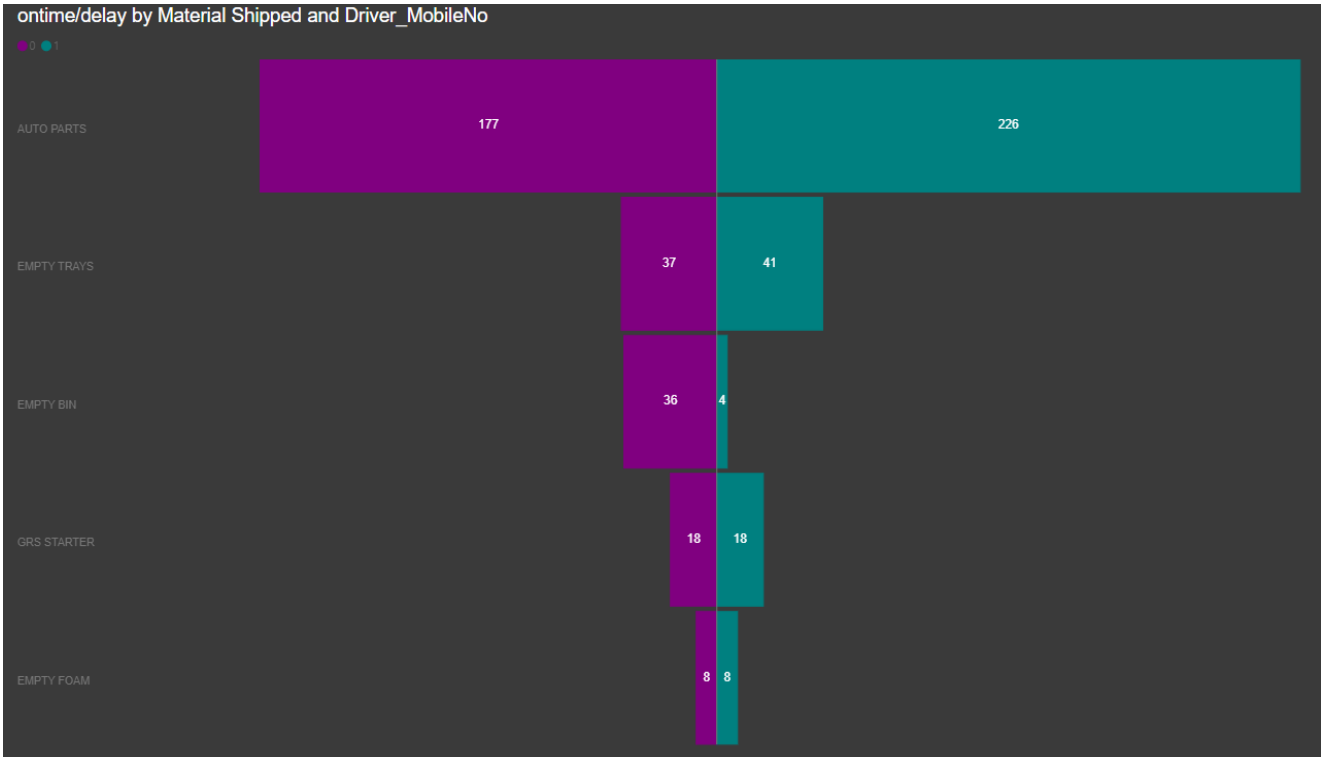
- In the line and clustered column chart below, we can see the total distance covered by each truck type with respect to how much of it was used in performing deliveries with relatively.
- The 32 FT Multi Axie 14 MT – HCV has covered the greatest number of distance by any other truck, which shows its resilience and performance reliability, as it is more widely used as well. Similarly, the lighter version of the same truck follows, 32 FT Single Axie 7 MT – HCV, which has covered the second highest distance, ensuring that this brand of truck is reputable and reliable for longer trips.
- Whilst it may have covered the most distance, the amount of deliveries that is performed by them is astonishingly low. This would bring us to a conclusion that this model of truck(s) is rather used for completing long distance deliveries, since it might be sturdier than the rest.



- One anomaly in this dataset is the “unknown”, in both the cases of vehicle type and driver name. Since the data for them was unavailable, it would be unwise to include them in our analysis. However, we can infer from rest of the data available that 32 FT Multi Axie 14 MT – HCV has covered the most number of deliveries, timely ones too. In fact, it has completed 100% of its deliveries, followed by 40 FT 3XL Trailer 35MT which has also completed 100% of its deliveries on time.



- In the treemap above, we can see the overall distance covered is done mostly by Haryana trucks, followed by Tamil Nadu trucks, which are slightly lower, but have far more deliveries completed than that of Haryana's.



- In this tornado chart, we can see the amount of deliveries completed with respect to the items that were shipped. Additionally, we want to verify whether a driver having a mobile number can have an impact on delivery.
- We can see that more deliveries for auto parts were completed with driver having a mobile phone. However, the rest do not really follow that pattern, and this goes to show that a driver carrying a mobile phone has little to do with the success of the delivery.
- While that is the overall picture, we cannot even merely paint that as conclusion of this story. There are several factors to consider in this scenario, for example the material that is to be shipped.
- If the material is to be deemed important, then the truck driver very likely will carry a mobile phone with him, in case of any emergency. In this chart, auto parts are relatively far more important than the other items, therefore more care was taken for materials of higher importance.