EDA OF CHARGING STATIONS DATASET

By

Saad Ur Rahman

Data Cleaning and Preparation

The data at hand is about 16 different charging stations over the course of 13 months. The initial dataset has 277 rows and 6 columns, which are ‘Start Time’, ‘Meter Start(Wh)’, ‘Meter End(Wh)’, ‘Total Duration’, and ‘charging station_name’. The columns ‘Meter Start(Wh)’ and ‘Total Duration’ are for integer type(integer values), columns ‘Meter End(Wh)’ and ‘Meter Total(Wh)’ are of float type(had decimal values), and columns ‘Start Time’ and ‘Charger_name’ are of object type(datetime/string/words).

The column ‘charging station_name’ had 13 missing values which I removed from the main dataset and stored as a separate data.

To ensure that charging data is logical, the rows with zero time consumption and positive power consumptions were initially removed and stored as a separate dataset. This ensured that no data implied that charging occured instantaneously, without any time lapse, which is impossible.

After the initial cleaning of the dataset, few graphs were plotted to check for basic trends and outliers, as shown below,

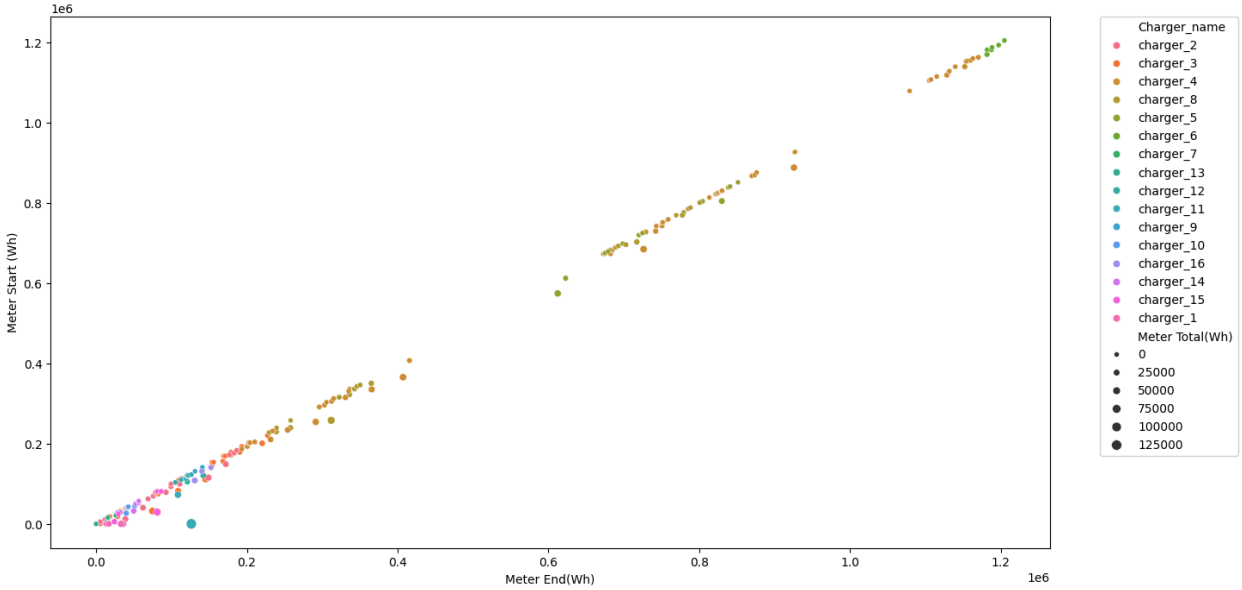

Figure 1: Scatter plot of Meter Start vs Meter End

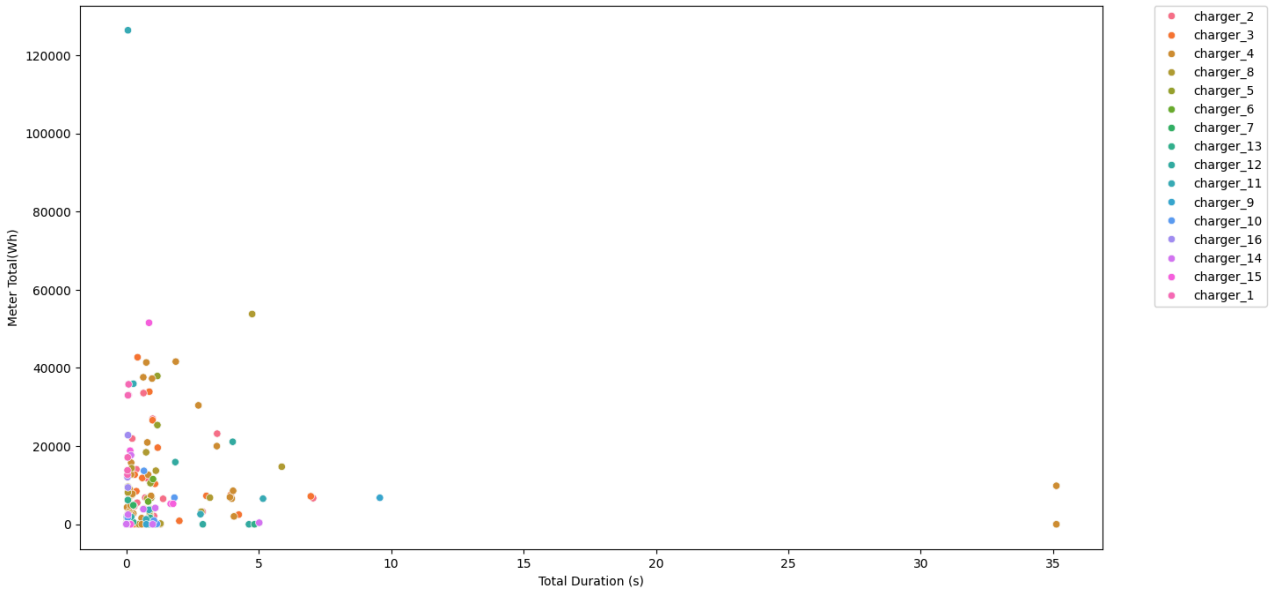

Figure 2: Scatter plot of Meter Total vs Time Duration

From the plots above we see that the charging stations follow a specific pattern, as the plot between the Meter Start and Meter End readings follow a straight line, with a gap aroud the 500000Wh and 120000Wh Mark, which we might be able to explained if further information is available.

Also, there are some outliers that have very high energy consumption or a very long duration. There were 4 such datapoint, which were removed from the dataset for more accurate and unbiased analysis.

Then the the average charging spead for each charging station was calculated and used to correct the Time Duration value in initially separated dataset by dividing Meter Total values with the average charging speed of the respective charging stations and these rows were added back to the main dataset.

Time Data Adjustment

The time data provided in the 'Start Time' column was of the format '31.08.2018 06:45', having two components, Date and Time(24 hours pattern). This was split into date, year, month, day, time, hour, minute, and year_month columns. These columns were then used to arrange the dataset in an ascending order with the earliest date and time on the top and the latest date and time at the bottom.

Statistical Summary

Some basic statistical analysis was performed and the findings are listed below.

Start Time column had 252 unique values
Meter Start (Wh) column had 225 unique values
Meter End(Wh) column had 249 unique values
Meter Total(Wh) column had 187 unique values
Total Duration (s) column had 211 unique values
charging station_name column had 16 unique values
date column had 117 unique values

year column had 2 unique values
month column had 11 unique values
day column had 31 unique values
time column had 214 unique values
hour column had 21 unique values
minute column had 60 unique values
year_month column had 13 unique values

The units of energy for ‘Meter Start (Wh)’, ‘ Meter End(Wh)’, ‘ Meter Total(Wh)’, columns were changed from Watt to KiloWatt and unit of time for ‘Total Duration (s)’ column was changed from seconds to hour. This was done to reflect the industry standards, which it Kilowatt Hour.

The mean and median for Energy Consumption(Meter Total) were 5.93 kW and 1.96 kW respectively.
The mean and median Time Time Duration were 17.41 Hours and 2.74 Hours respectively.

Charging Station-specific Analysis

After the aforementioned changes were made, charghing station specific EDA was performed.

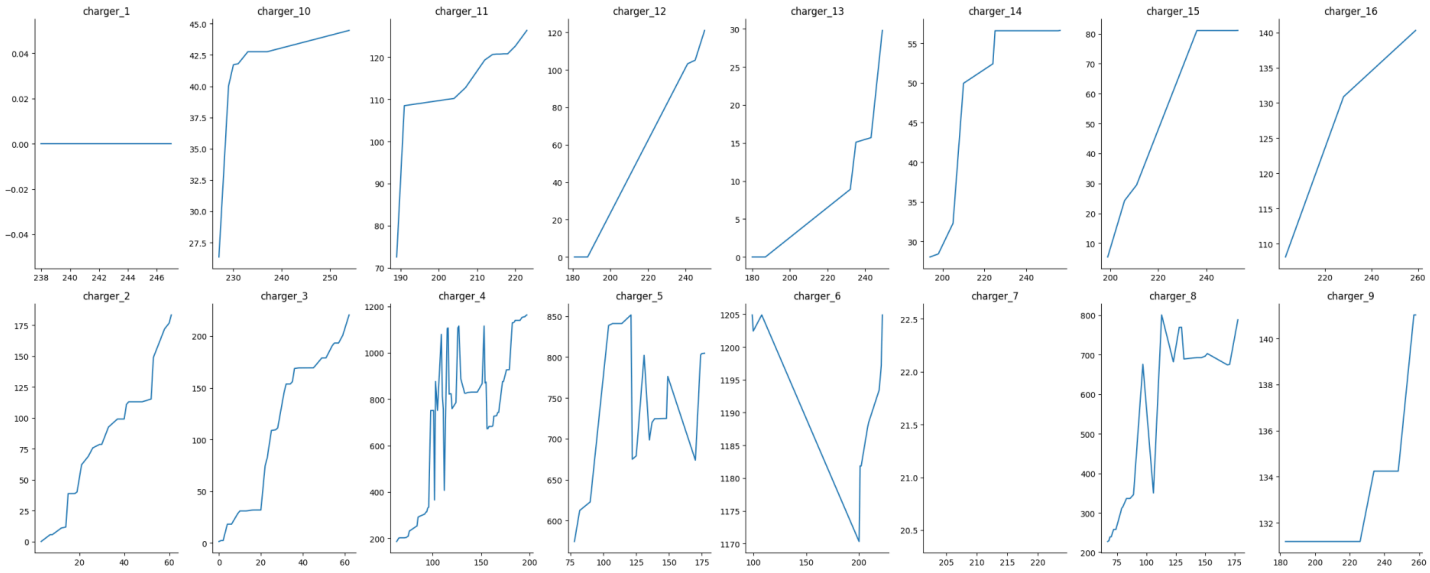

Figure 3: Initial Meter readings per instance for each charging station

From the plot above, we can see that the most of the charging station followed a continuous meter reading format, where the meter readings were not reset after each use, this is the usually basic standard. However there were some expectations, with charging station 1’s meter was reset after every use, charging station 7 was only used once, and charging stations 4, 5, and 8 showed fluctuations in the initial Meter readings, this could be due to malfunction of the meter or manual intervention.

Looking at the number of times a charging station was used, as shown in the figure below,

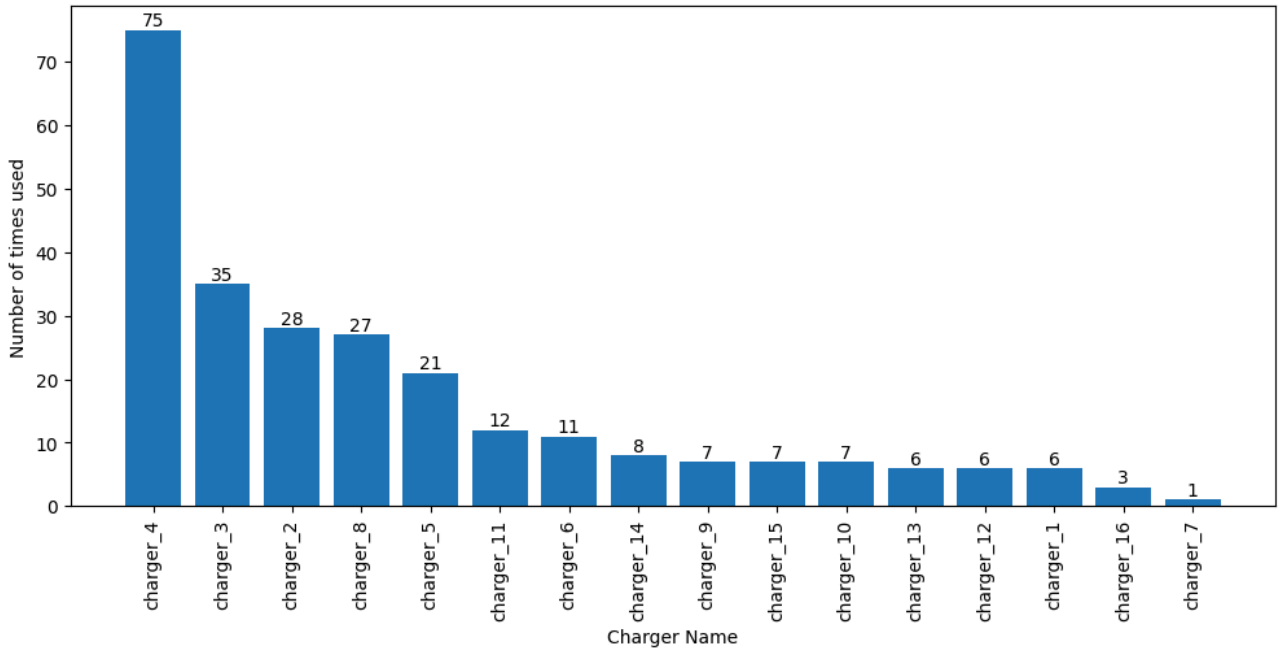

Figure 4: Number to chargings per charging station

It can be seen that, charging station 4 was used most number of times followed by charging station 3, 2, 8, 5, and so on, while charging station 7 was used only once. This can be due to placement, location, and accessibility of the charging stations.

This usage is also reflected in the total energy consumption per charging station, as show below,

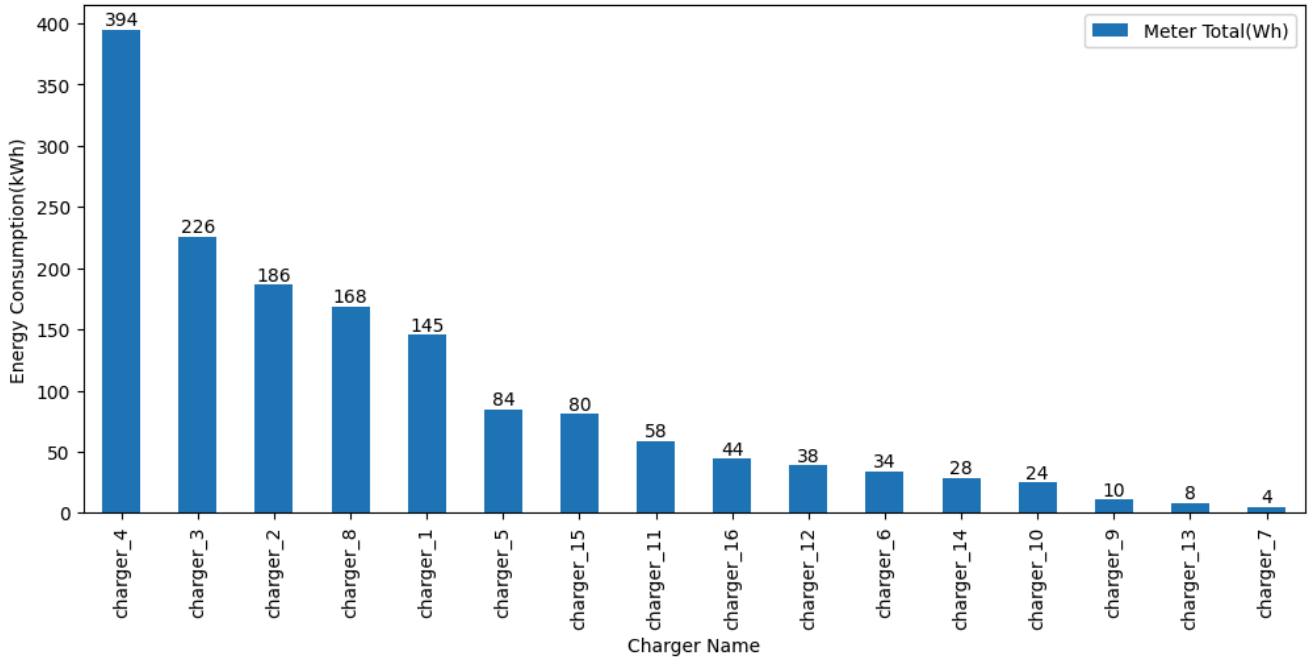

Figure 5: Total Energy Consumption per Charging Station

Charging station 4 consumed most amount energy of about 394 kiloWatt followed by charging stations 3, 2 and 8 while charging station 7 consumed least amount of energy, with charging station 4 of about 4 kiloWatt. However, there are few execeptios such as charging stations 1 and 16, which consumed much more energy than expected.

These exceptions can be explained by the figure below,

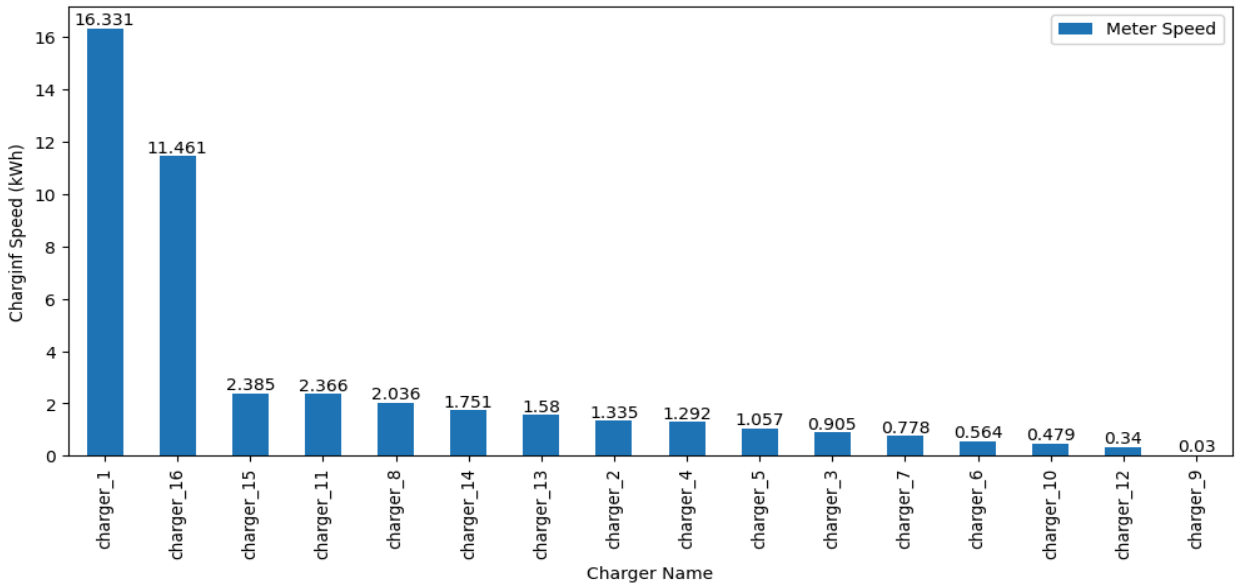

Figure 6: Charging Speed/Energy Consumption of each Charging Station

It can be observed that charging stations 1 and 16 have much higher speed or energy consumption, as compared to other charging stations, hence they consume much more energy in comparatively less amount of time or number of instances.

Temporal Analysis

Time-based Analysis was performed to further understand the underlying charging trends.

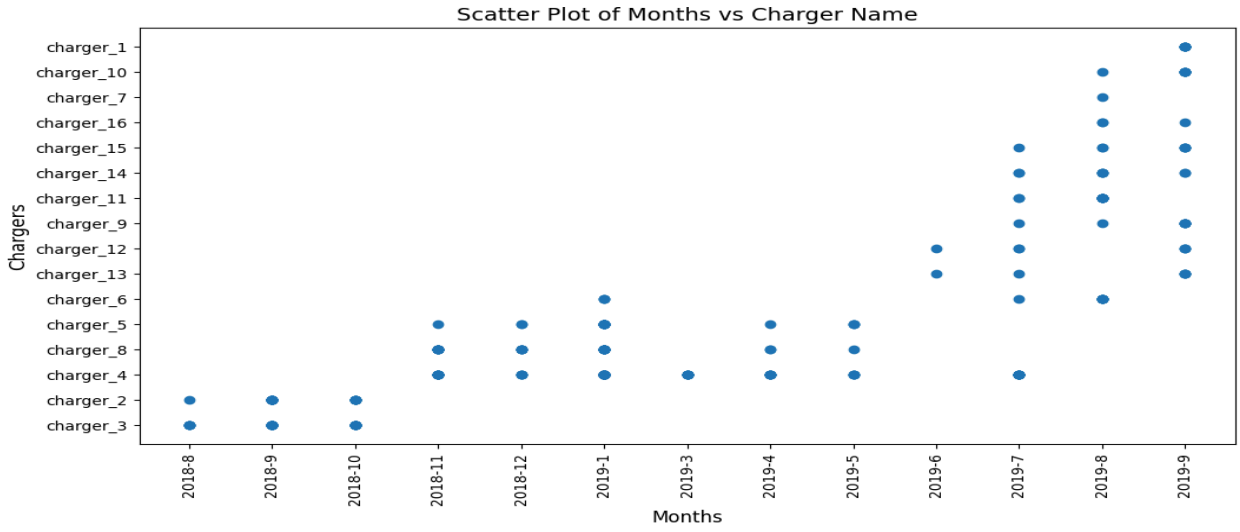

Figure 7: Scatter Plot of Charging Staions vs Months

In scatter plot above, it can be observed that there is a very specific pattern, with charging stations 2 and 3 being used for the first 3 months, followed 4, 5, and 8 for the next 6 months, with the exception of March and July 2019, and rest of the charging stations being used for the last 3 to 4 months.

We can infer from this trend that the charging stations were either being used for a time and then replaced by others or were just put out of service. Given the knowledge we have in hand, it is hard to conclude which one of the two is the case. However, looking at the charging speed of individual charging stations as seen in Figure 6 , the charging stations used in latter months such as 1, 15, 16, 11 have higher charging speed than the ones used earlier, implying that the former hypothesis might be true.

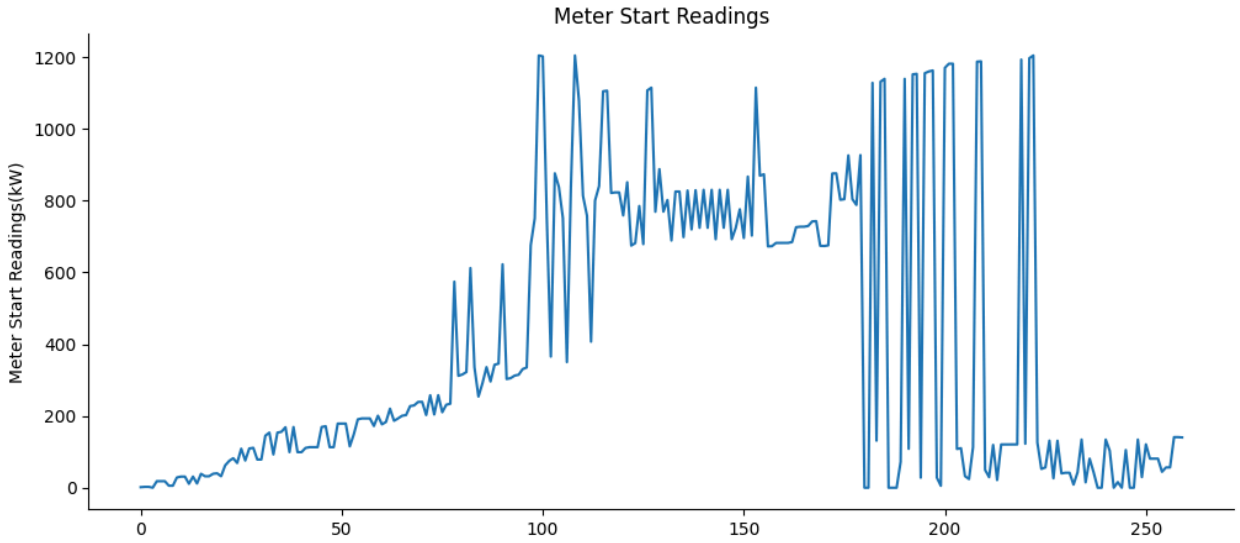

Figure 8: Meter Start Readings over the period of time

In the above plot of the meter start reading vs time(dataset was arranged based on time), we can different patters over the course of time see a gradual increase in meter start readings, with some fluctuations and plateauing in the middle followed by high fluctuation in the latter part. This trend can be explained by the Figure 3 and Figure 7.

The charging stations used during the initial months generally followed a continuous meter reading format, therefore resulting in a gradual increase in the overall meter reading during that time, which was followed by fluctuations and plateauing(with fluctuations), which might be due to charging stations 4, 5, 6, and 8. The major fluctuations at the latter dates might be due to the use of charging stations 4, 6, 9 and 13 which have a wide range of meter start dates among each other. The final small fluctuations might be due to charging station 1 and others since the Start Meter readings for charging station 1 were always zero.

Furthemore, the total energy consumption readings alsofollow an interesting pattern as depicted in the figure below,

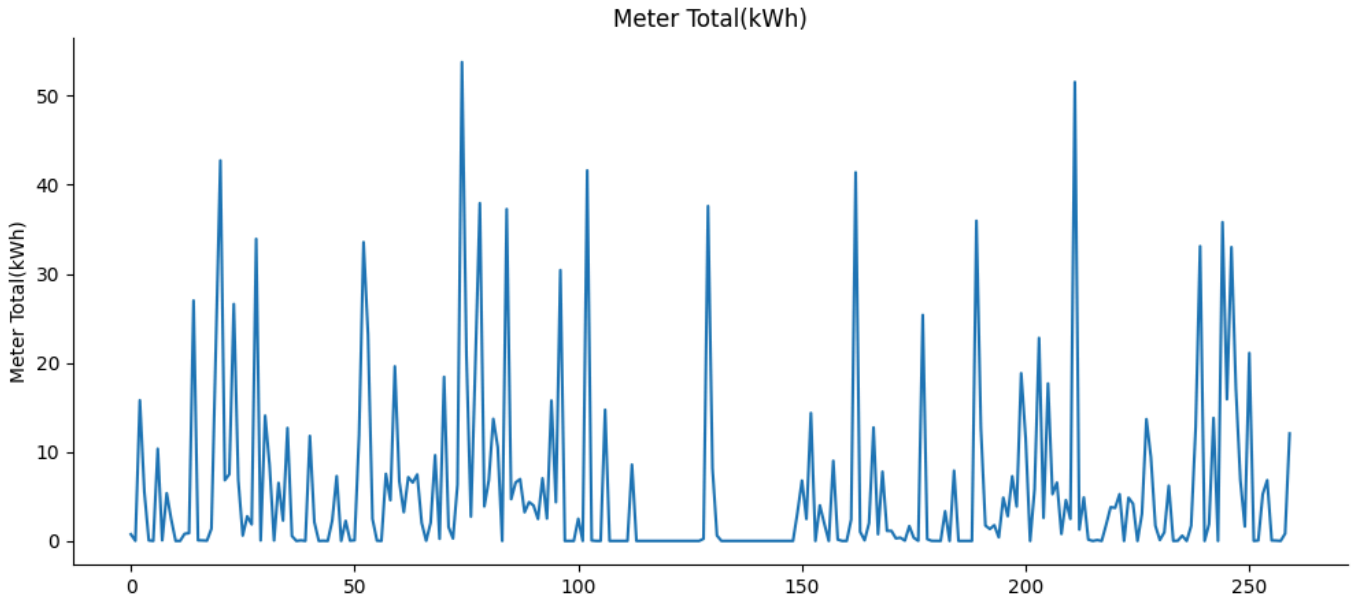

Figure 9: Energy Consumption over the period of time

There is a rough pattern visible in the graph above, with periodic spikes of energy consumption being recorded, apart from a section in the middle where the energy consumption usually remains close to zero, with just one major reading. This can be better understood by the energy consumption graph depicted below,

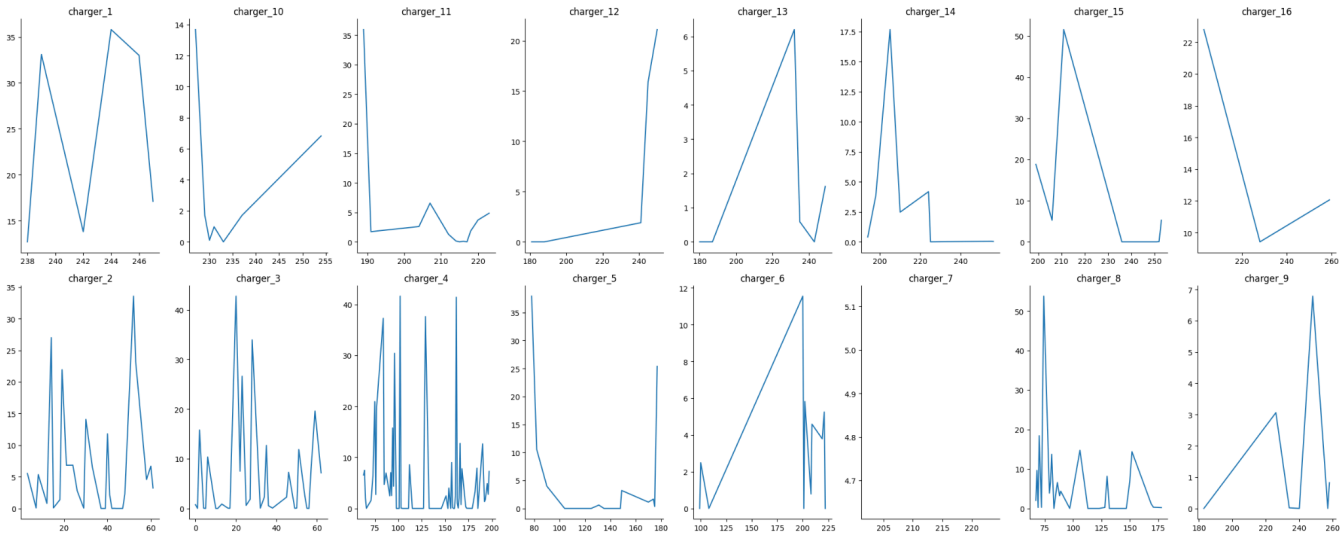

Figure 10:Energy Consumption over the period of time for each Charging Station

As seen in figure above, only charging station 4 consumed a significant amount of energy near the 125 mark on x-axis, which coincides with the total energy consumption graph plotted previously.

We might be able to understand the above shown energy consumption pattern better by using temporal analysis models such as ARIMA/SARIMA. There might be some seasonal patterns that can be understood and used for further data analysis and prediction models.

Moreover, plotting graphs against ‘hours’ showed some interesting charging patterns as well,

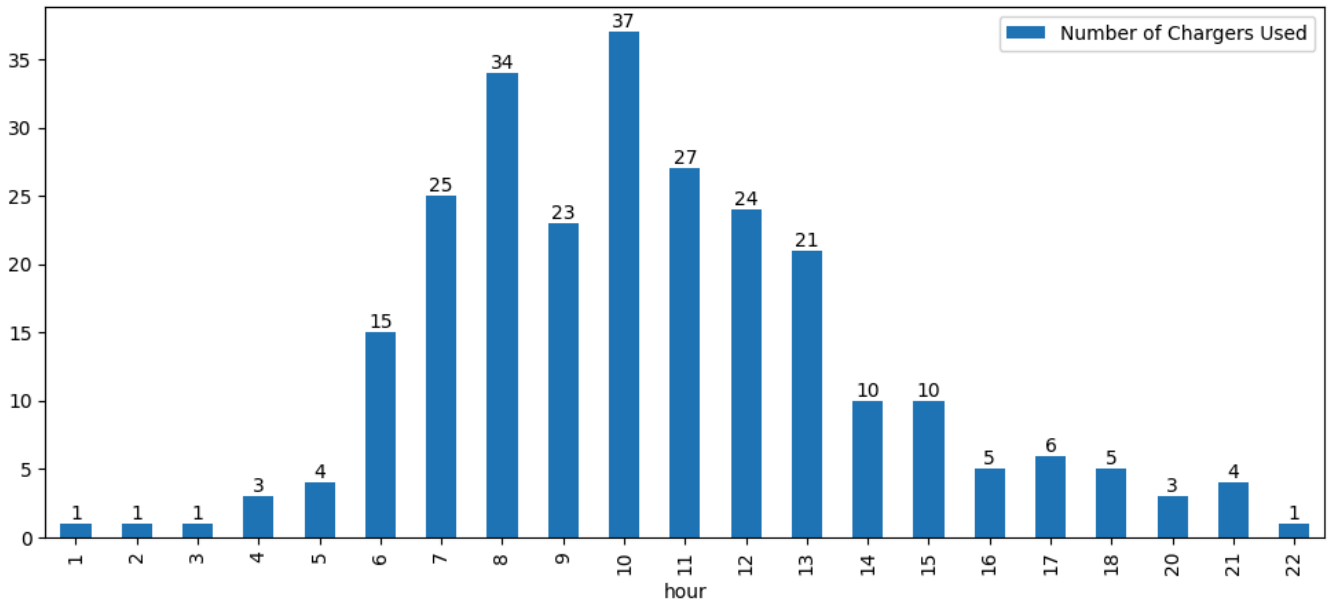

Figure 10: Number of times charging stations were used vs hours of the day

In the figure above, the graph roughly follows a bell curve, with the number of charging stations used rising at 6 AM in the morning, peaking at 10 AM, with 37 times a charging station being used at this hour and then dropping to just 5 after 3 PM. However, we can also see a drop at 9 AM, with comparatively less number of times charging stations being used as compared to 8 AM and 10 AM.

We can infer that this probably follows a standard human lifestyle, with people using charging stations after waking up in the morning, with a drop at 9 AM, which is usually the start of office/business hours.

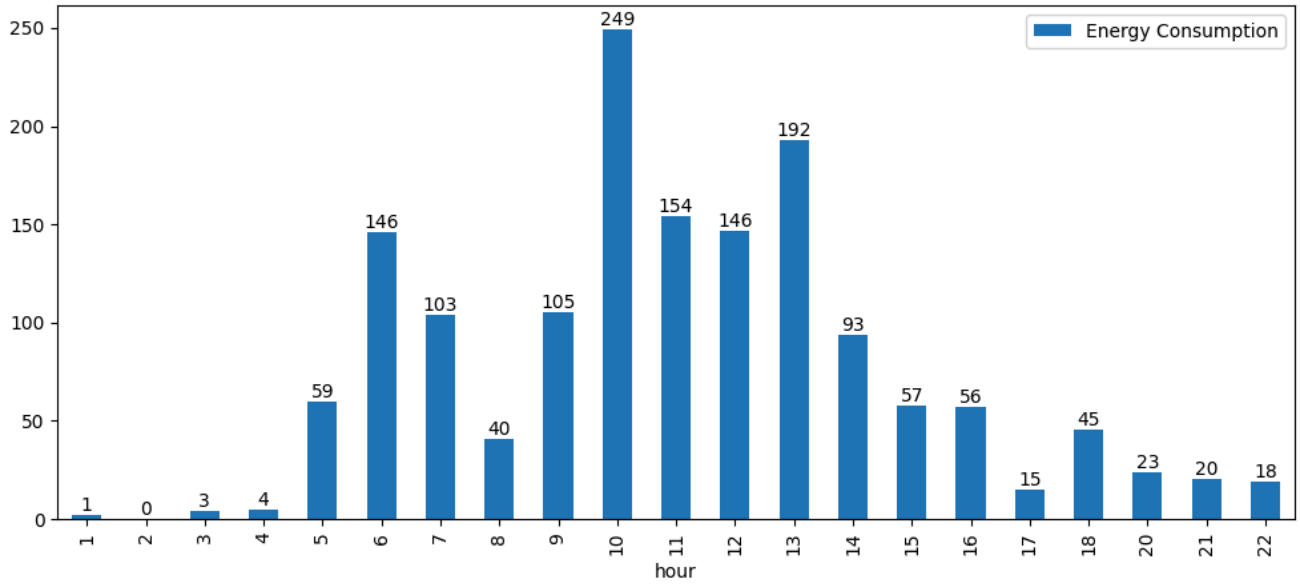

Figure 11: Total Energy Consumed vs hours of the day

The energy consumption graph in Figure 11 also follows a somewhat bell curve in the figure above.This can directly be related to the number of charging stations that were used per hour. However, we can observe more fluctuations as compared to Figure 10, which can be due to the different charging speeds of individual charging stations being used.

Distribution Analysis

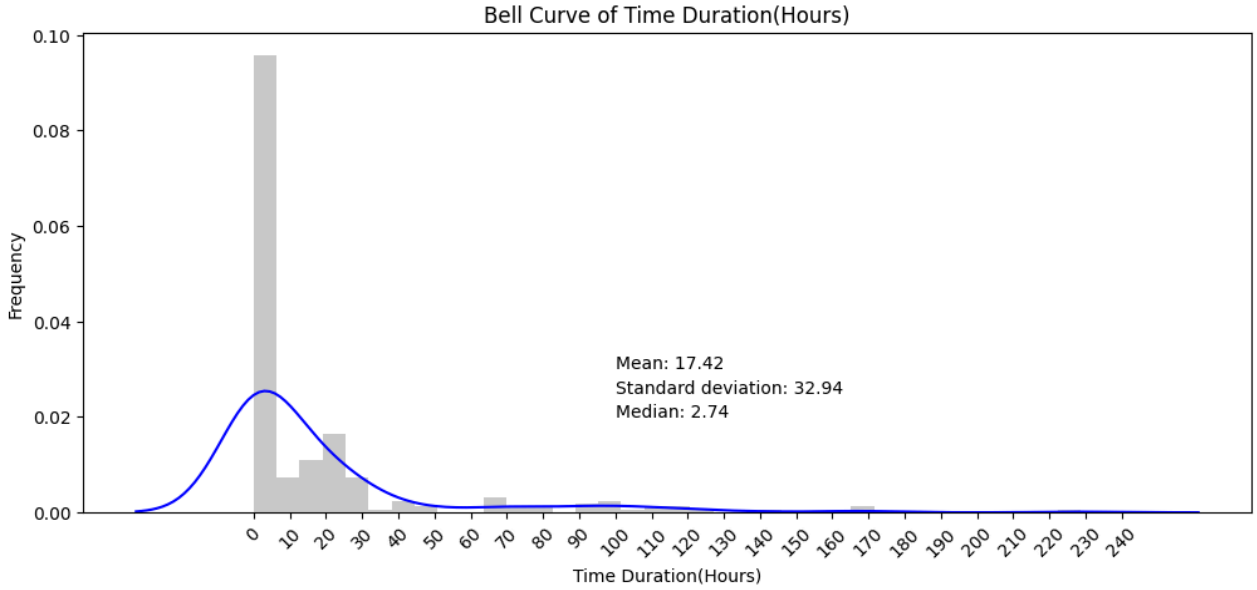

Figure 12: Distribution plot of Time Duration of charging stations(hours)

As observed from the distribution graph above, the majority of chargings lasted less than 10 hours, with few going up to 20 to 30 hours, while very few lasting more than that.

Although the average time taken is around 17.5 hours, this is due to the few charging instances that went on for a much longer time. This can be understood by looking at the median charging time, which is under 2.74, stating that 50% of the charging instances did not last more than 2 hours 45 minutes.

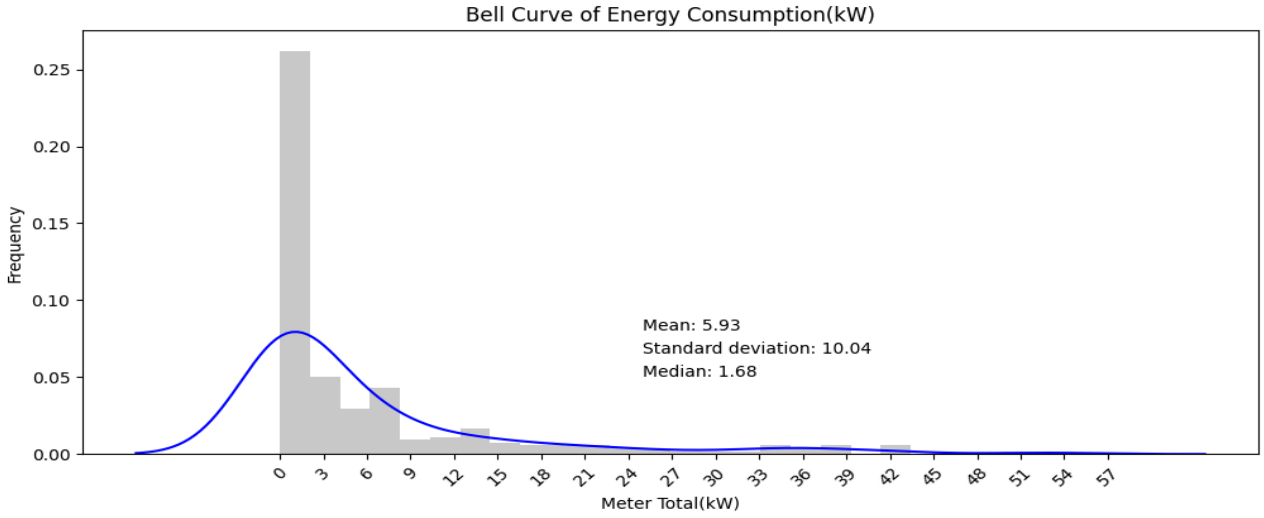

Figure 13:Distribution plot of Energy Consumption of charging stations

Similarly in the graph above, we can see that the majority of the energy consumption per charging session was below 3 kilowatt. Furthermore, as calculated and displayed, the mean was around 6 kW, which is again is due to a few instances where the energy was consumed up to 42kW and more. Meanwhile, the median stood at 1.68, which implies that 50% of the charging instances consumed energy less than 1.68 kW.

Conclusion

- Most of the stations followed linear patterns in meter readings, with a consisten rise in readings.
- Some stations charged faster than others, these stations were used few months later than the slower ones.
- Some stations were used much more number of times than others, potentially skewing the dataset
- Observed peak charging times in during morning and noon hours with declines around evening..
- 50% of the charging instances did not last more than 2 hours 45 minutes
- 50% of the charging instances consumed energy less than 1.68 kW.
- Enhancements can be made for predictive maintenance and charger deployment based on usage patterns.

Future Suggestions

With the help of additional data, we can find more intrinsic patterns in charging times and frequencies, which can help build predictive models that could be developed to forecast potential downtimes and maintenance needs.

Furthermore, Insights from hourly and charging station-specific analyses could guide the strategic placement and scheduling of charging stations to maximize efficiency and user satisfaction.

Finally, specific data can be procured to understand the unerperformance of charging stations, resulting in more compatible design and construction.