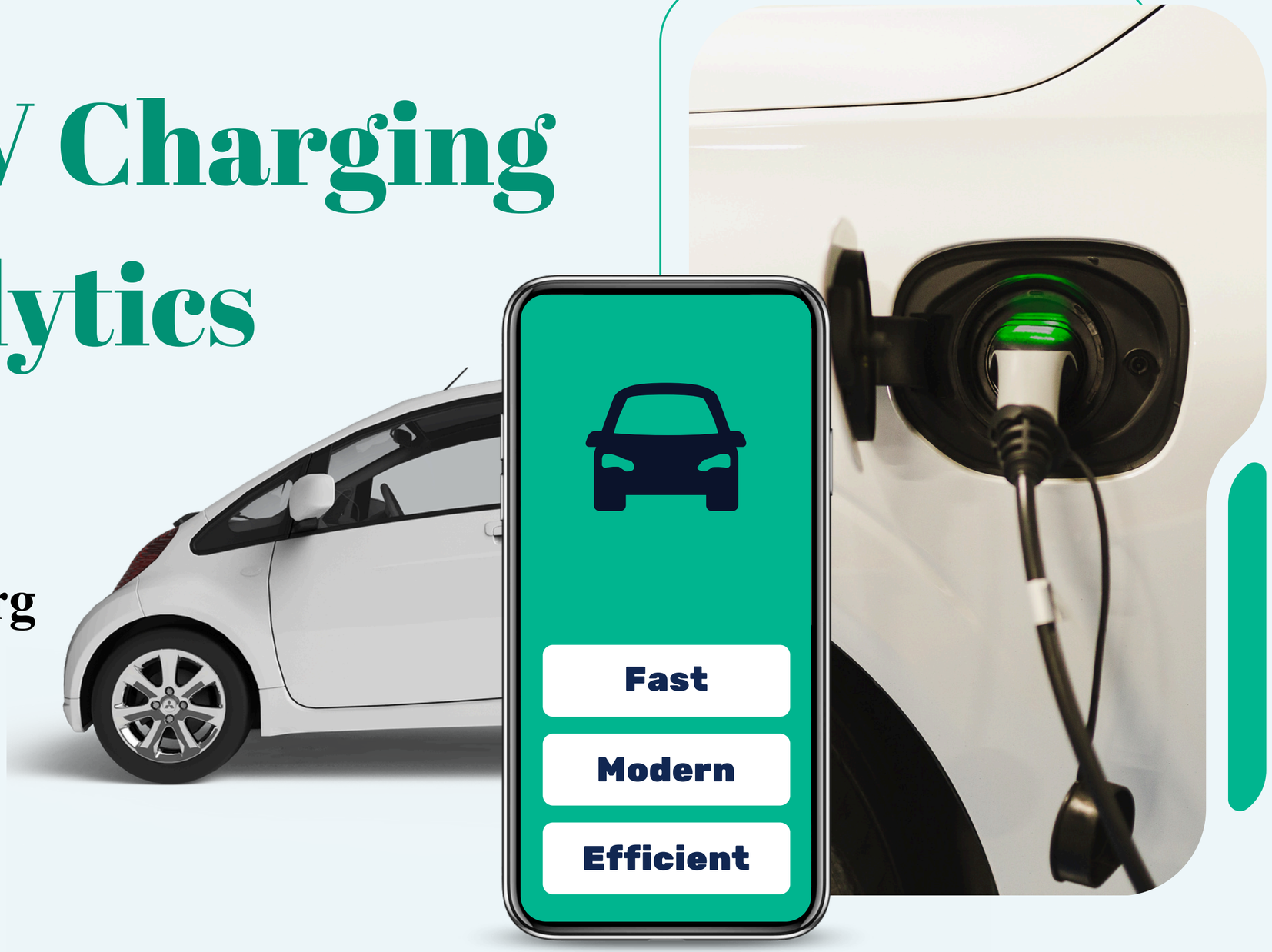
Warsaw EV Charging Analytics

Presented by: Rashi Garg



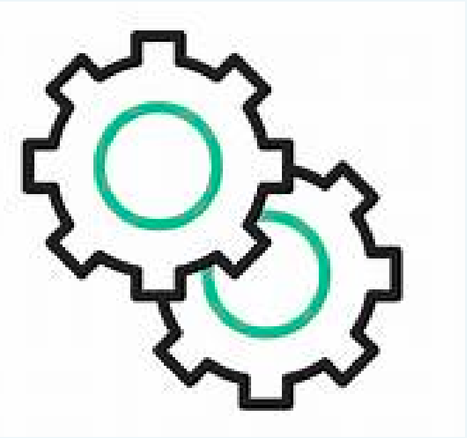
Project Introduction

This project involves a comprehensive data analysis of the Electric Vehicle (EV) charging infrastructure in Warsaw. Working with datasets covering city districts, charging stations, customer demographics, and over a year of charging sessions, the primary goal was to evaluate infrastructure readiness and optimize expansion planning.

Key areas of focus for the analysis include:

- Demand Mapping: Identifying high-demand charging locations and districts where current infrastructure is struggling to keep up with usage.
- Customer Behavior: Analyzing charging patterns and session durations across different income tiers to understand how different demographics utilize the network.
- Infrastructure Growth: Assessing whether the current number and type of charging stations can support the city's projected EV growth.
- Strategic Recommendations: Providing data-driven insights to help city planners determine where to prioritize new installations for maximum efficiency and accessibility.

Project Execution Summary



1. SQL

- Data Engineering: Cleaned and joined 4 core datasets.
- Diagnostic Analytics: Resolved 14 business queries to identify infrastructure gaps.

2. Power BI

- Interactive Reporting: Developed a 4-page dashboard for geospatial and behavioral insights.
- KPI Tracking: Visualized demand, revenue, and customer segments.

3. Machine Learning

- Predictive Modeling: Forecasted future charging demand for urban planning.
- Optimization: Used Regression and Clustering for cost and user segmentation.

PHASE -1

SQL ANALYSIS

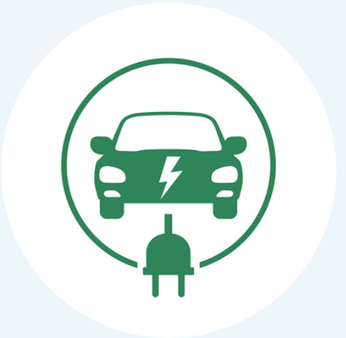
1. List all districts along with their income tiers.

```sql
# 1. list all districts along with their income tiers.
select district_name,income_tier from districts;
```

Result

district_name	income_tier
Wilanów	High
Śródmieście	High
Żoliborz	High
Mokotów	Mid-Range
Wola	Mid-Range
Ursynów	Mid-Range
Bielany	Mid-Range
Praga Południe	Low-Mid
Ochota	Mid-Range

2. Count the total number of charging stations in each district.



```sql
# 2. Count the total number of charging stations in each district.
select district_name,count(station_id) as total_stations from charging_stations
group by district_name;
```

Result

district_name	total_stations
Ursynów	12
Włochy	6
Wilanów	8
Wesoła	11
Ochota	5
Ursus	2
Żoliborz	2
Bemowo	2
Mokotów	5

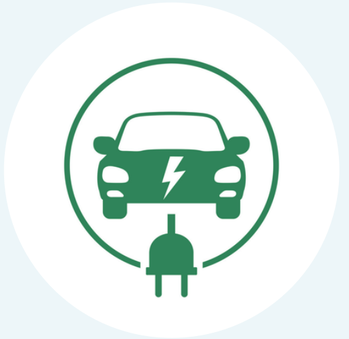
3. Find the total number of charging plugs available per district.

```sql
# 3. Find the total number of charging plugs available per district.
select district_name,sum(plugs_count) as total_available_plugs
from charging_stations
group by district_name;
```

Result

district_name	total_available_plugs
Ursynów	72
Włochy	36
Wilanów	48
Wesoła	66
Ochota	30
Ursus	12
Żoliborz	12
Bemowo	12
Mokotów	30

4. Calculate the total number of charging sessions recorded.

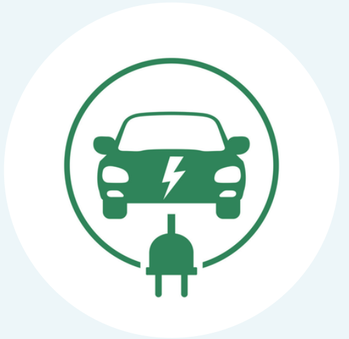


```sql
# 4. Calculate the total number of charging sessions recorded.
select count(*) as total_sessions
from charging_sessions;
```

Result

total_sessions
114000

5. Find the average energy charged (kwh) per charging sessions .

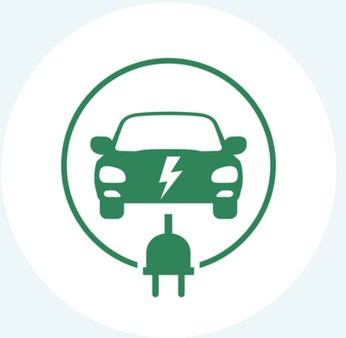


```sql
# 5. Find the average energy charged (kWh) per charging session.
select session_id,round(avg(kwh_charged),2) as average_energy_kwh from charging_sessions
group by session_id;
```

Result

session_id	average_energy_kwh
S00001	46.69
S00002	39.06
S00003	30.28
S00004	44.58
S00005	38
S00006	42.52
S00007	47.13
S00008	20.68
S00009	26.96

6. Calculate total energy charged (kwh) by each district.

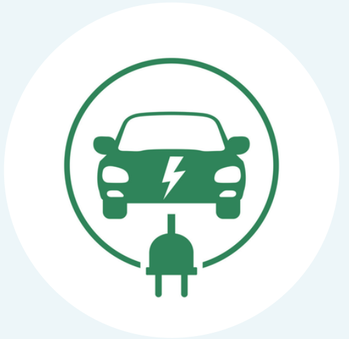


```sql
# 6. Calculate total energy charged (kWh) by each district.
select districts.district_name,round(sum(charging_sessions.kwh_charged),2)
as total_energy from charging_sessions
join charging_stations
using(station_id)
join districts
using(district_name)
group by districts.district_name;
```

Result

district_name	total_energy
Wesoła	767641.26
Wola	139796.55
Ochota	339168.45
Mokotów	339212.79
Włochy	409485.95
Żoliborz	141089.31
Ursynów	820629.82
Śródmieście	205791.69
Targówek	341296.66

7. Identify the top 5 charging stations based on number of charging sessions .



```sql
# 7. Identify the top 5 charging stations based on number of charging sessions.
select station_id,count(session_id) as total_sessions from charging_sessions
group by station_id
order by total_sessions desc
limit 5;
```

Result

station_id	total_sessions
163	1856
150	1851
27	1825
92	1824
68	1811

8 .Calculate total charging revenue generated per district.

```sql
# 8. Calculate total charging revenue generated per district.
select charging_stations.district_name,
round(sum(charging_sessions.total_cost),2) as total_revenue
from charging_sessions
join charging_stations
using(station_id)
group by charging_stations.district_name;
```

Result

district_name	total_revenue
Wesoła	1964951.35
Wola	357333.01
Ochota	868531.89
Mokotów	871968.22
Włochy	1051339.58
Żoliborz	359931.26
Ursynów	2101299.59
Śródmieście	525634.52
Targówek	876870.2

9. Find the average charging cost per session for each income tier .

```sql
# 9. Find the average charging cost per session for each income tier.
select customers.income_tier,round(avg(total_cost),2) as average_charging_cost
from charging_sessions
join customers
using(customer_id)
group by customers.income_tier;
```

Result

income_tier	average_charging_cost
Mid-Range	117.78
Low-Mid	75.19
High	142.45

10. Analyze average energy charged per customer.

```sql
# 10. Analyze average energy charged per customer.
 select customers.customer_id,
 round(avg(charging_sessions.kwh_charged),2)
 as average_energy
 from customers join charging_sessions
 using(customer_id)
 group by customers.customer_id;
```

Result

customer_id	average_energy
C3806	49.27
C4931	42.39
C2250	27.7
C3790	48.18
C4834	31.5
C6834	48.89
C0967	43.25
C5713	17.06
C5256	32.93

11 . Compare projected EV count with total available charging plugs for each district .

```sql
 # 11. Compare projected EV count with total available charging plugs for each district.
select districts.district_name,districts.projected_evs,
round(sum(charging_stations.plugs_count),2) as available_plugs,
(districts.projected_evs-sum(charging_stations.plugs_count)) as gap,
case when districts.projected_evs>sum(charging_stations.plugs_count)
then'shortage'
when districts.projected_evs<sum(charging_stations.plugs_count)
then 'sufficient'
else 'balanced'
end as infrastrucrure_status
from districts join charging_stations
using (district_name)
group by  districts.district_name,districts.projected_evs;
```

Result

district_name	projected_evs	available_plugs	gap	infrastrucrure_status
Ursynów	2900	72	2828	shortage
Włochy	1200	36	1164	shortage
Wilanów	3100	48	3052	shortage
Wesoła	700	66	634	shortage
Ochota	1600	30	1570	shortage
Ursus	1100	12	1088	shortage
Żoliborz	1800	12	1788	shortage
Bemowo	1800	12	1788	shortage
Mokotów	3500	30	3470	shortage

12 . Identify districts where projected EVs exceed available charging plugs .

```sql
# 12. Identify districts where projected EVs exceed available charging plugs.
select districts.district_name,districts.projected_evs,
round(sum(charging_stations.plugs_count),2) as available_plugs,
(districts.projected_evs-sum(charging_stations.plugs_count)) as gap
from districts join charging_stations
using (district_name)
group by  districts.district_name,districts.projected_evs
having districts.projected_evs>sum(charging_stations.plugs_count);
```

Result

district_name	projected_evs	available_plugs	gap
Ursynów	2900	72	2828
Włochy	1200	36	1164
Wilanów	3100	48	3052
Wesoła	700	66	634
Ochota	1600	30	1570
Ursus	1100	12	1088
Żoliborz	1800	12	1788
Bemowo	1800	12	1788
Mokotów	3500	30	3470

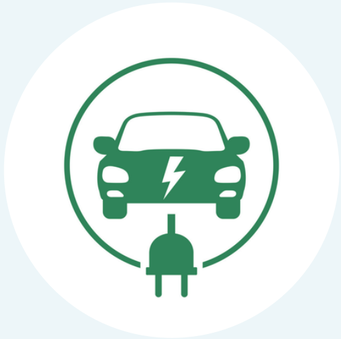
13. Analyze charging demand per projected EV by districts (total Kwh/projected EVs) .

```sql
# 13. Analyze charging demand per projected EV by district (total kWh / projected EVs).
select districts.district_name,districts.projected_evs,round(sum(charging_sessions.kwh_charged),2)
as total_charging_demand,
round((sum(charging_sessions.kwh_charged)/districts.projected_evs),2)
as demand_per_projected_ev ,case when districts.projected_evs>sum(charging_sessions.kwh_charged) then 'low'
else 'high'
end as demand_status
from charging_stations
join charging_sessions
using (station_id)
join districts
using(district_name)
group by districts.district_name,districts.projected_evs;
```

Result

district_name	projected_evs	total_charging_demand	demand_per_projected_ev	demand_status
Wesoła	700	767641.26	1096.63	high
Wola	3300	139796.55	42.36	high
Ochota	1600	339168.45	211.98	high
Mokotów	3500	339212.79	96.92	high
Włochy	1200	409485.95	341.24	high
Żoliborz	1800	141089.31	78.38	high
Ursynów	2900	820629.82	282.98	high
Śródmieście	2700	205791.69	76.22	high
Targówek	1400	341296.66	243.78	high

14. Rank districts based on infrastructure readiness using actual charging usage vs projections.



```sql
# 14. Rank districts based on infrastructure readiness using actual charging usage vs projections.
select districts.district_name,districts.projected_evs,round(sum(charging_sessions.kwh_charged),0) as actual_usage,
round(sum(charging_sessions.kwh_charged)/districts.projected_evs,2) as readiness_score ,
dense_rank() over (order by round(sum(charging_sessions.kwh_charged)/districts.projected_evs,2) DESC)
as readiness_rank
from charging_sessions join charging_stations
using (station_id)
join districts
using(district_name)
group by  districts.district_name,districts.projected_evs;
```

Result

district_name	projected_evs	actual_usage	readiness_score	readiness_rank
Wesoła	700	767641	1096.63	1
Włochy	1200	409486	341.24	2
Ursynów	2900	820630	282.98	3
Targówek	1400	341297	243.78	4
Rembertów	600	138615	231.03	5
Ochota	1600	339168	211.98	6
Wilanów	3100	546042	176.14	7
Ursus	1100	133624	121.48	8
Mokotów	3500	339213	96.92	9

Insights from SQL analysis

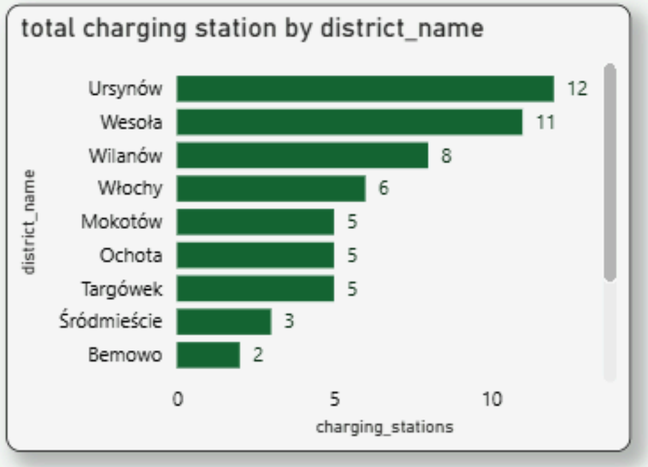
- Infrastructure Gap: EV charging stations are concentrated in only a few districts, leaving several areas in Warsaw underserved.
- Peak Demand: Charging activity spikes during the day and early evening, creating significant pressure on the grid during these hours.
- High-Usage Segments: Premium EV owners and high-income customers account for the largest share of total energy consumption per session.
- Revenue Concentration: A small number of districts and specific customer groups generate the majority of total charging revenue.
- Future Bottlenecks: Current infrastructure is not keeping pace with projected EV growth, which may lead to future congestion and system stress.

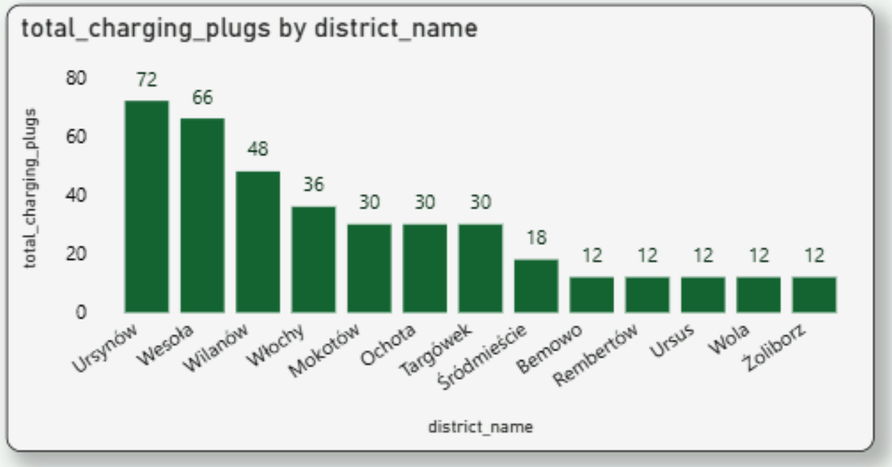
PHASE -2

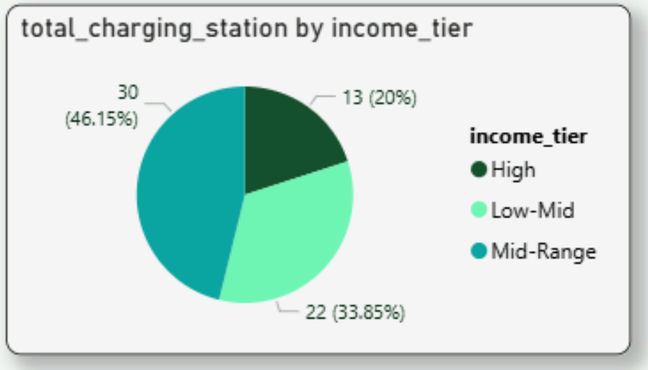
POWER BI ANALYSIS

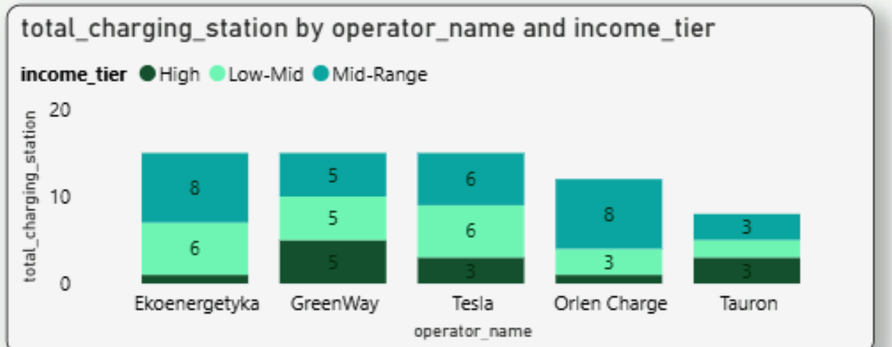
EV Infrastructure Overview

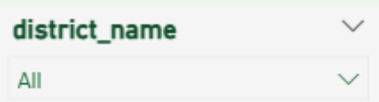
390	18	6.00	65
total_charging_plugs	total_districts	avg_plugin	total_charging_station

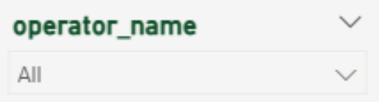










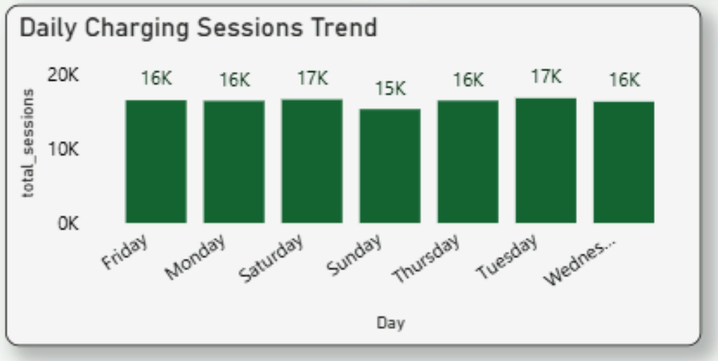


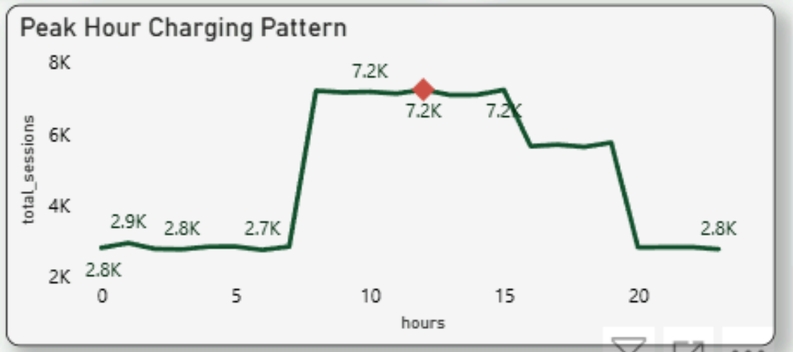
Insights

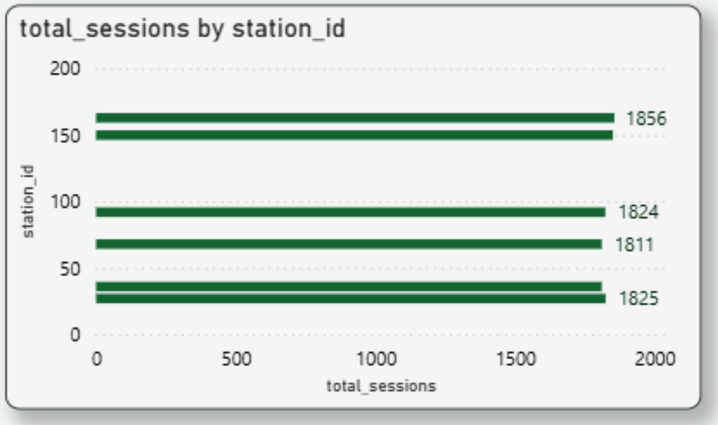
- Uneven Distribution: Charging stations are not balanced across the city; Ursynów (12) and Wilanów (11) have the highest density, while other districts remain under-served.
- Income & Usage Correlation: While Mid-Range districts hold the most stations (46%), the highest energy consumption and revenue come from High-Income customers and premium EV models.
- Peak Demand Pressure: Charging activity is most concentrated during daytime and early evening, creating significant load on the infrastructure during these hours.
- Capacity Overview: The network currently operates with 65 stations and 390 plugs, maintaining an average ratio of 6.00 plugs per station.
- Future Growth Risk: Current infrastructure is not aligned with projected EV growth, which may lead to congestion and system stress in the near future.

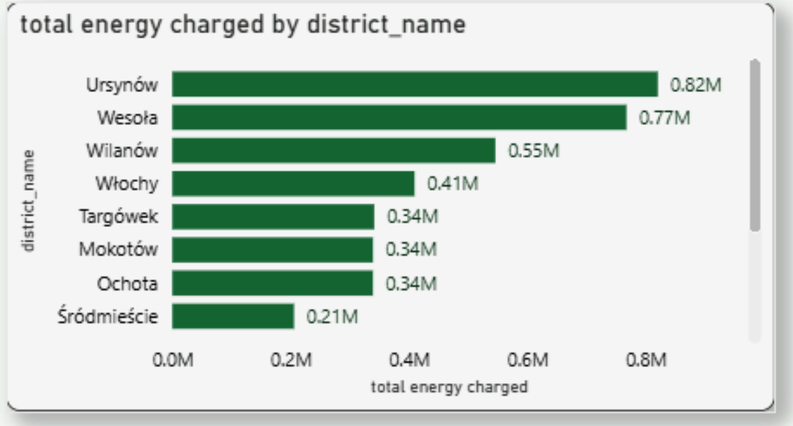
Charging Usage And Demand

114K	39.13	4.46M	12
total_sessions	average_energy_per_session	total_energy_charged(kwh)	Peak Hour









district_name

All

income_tier

- [ ] High
- [ ] Low-Mid
- [ ] Mid-Range

operator_name

All

Insights

- Geographic Gaps: Infrastructure is heavily concentrated in districts like Ursynów (12 stations) and Wilanów (11 stations), while areas like Bemowo and Bielany are underserved with only 2 stations each.
- High-Value Targets: Revenue is driven by a limited number of districts, with High-Income customers and premium EV models consuming significantly more energy per session.
- Peak Load Pressure: There is a major demand spike at 12:00 PM (Hour 12), with the heaviest system pressure occurring between 10:00 AM and 4:00 PM.
- Station Performance: While the network averages 6.00 plugs per station, high-traffic units like Station ID 1856 handle nearly 2,000 sessions, requiring prioritized maintenance.
- Growth Misalignment: Current station placement does not match projected EV growth in several districts, which will likely lead to future congestion.
- Operational Scale: The network has successfully managed 114K total sessions and delivered 4.46M kWh of energy, showing a strong and growing user base.

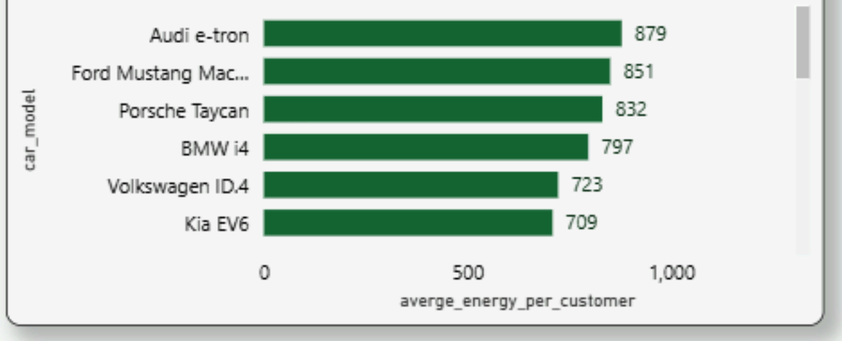
Customer Behaviour Analysis

8000
total_customers

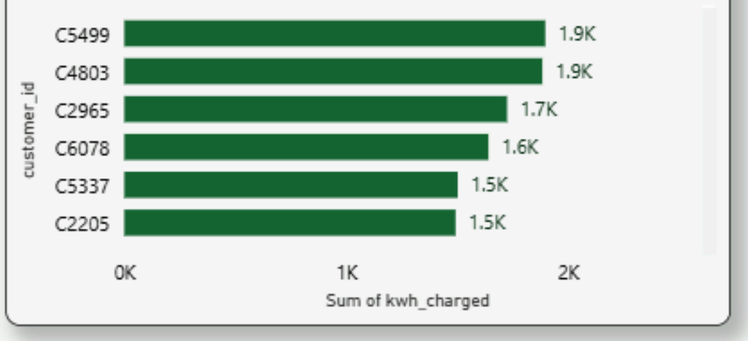
60.20
averge_battery_capacity

557.53
averge_energy_per_customer

Average Energy Consumption by car_model



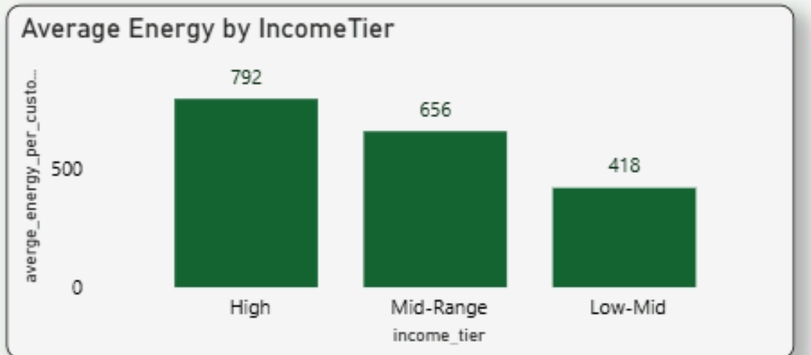
Top Energy Consuming Customers



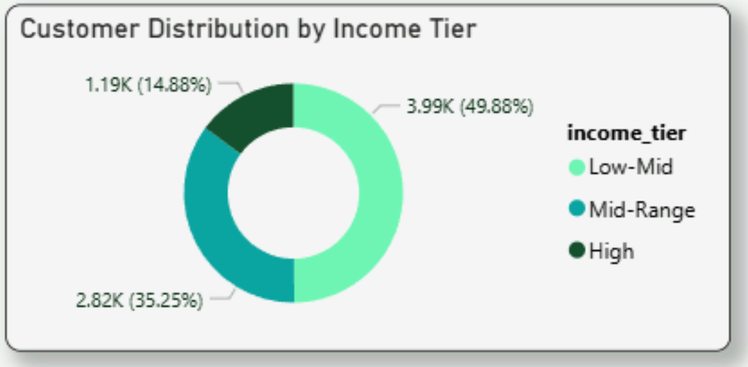
district_name

All

Average Energy by IncomeTier



Customer Distribution by Income Tier



income_tier

- ☐ High
- ☐ Low-Mid
- ☐ Mid-Range

operator_name

All

Insights

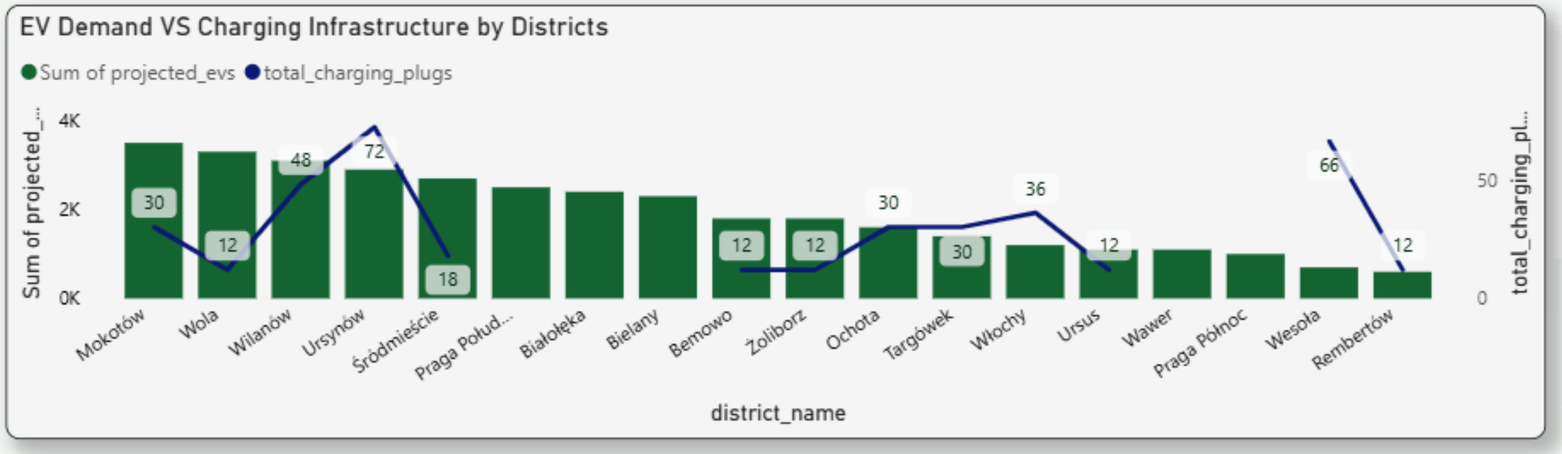
- Core Customer Base: The network currently serves a total of 8,000 customers.
- High-Energy Models: Premium EV models, such as the Audi e-tron and Ford Mustang Mach-E, lead in energy consumption per session.
- Income Tier Impact: High-income customers consume the most energy, averaging nearly 800 kWh, which is significantly higher than Mid-Range or Low-Mid tiers.
- Customer Distribution: The largest segment of the customer base (49.85%) belongs to the Mid-Range income tier.
- Individual Usage: On average, each customer consumes 557.53 kWh of energy.
- Battery Capacity: The average battery capacity across the analyzed fleet is 60.20 kWh.
- Top Consumers: A small group of power users (IDs like C5489 and C4065) are responsible for the highest individual energy loads, reaching up to 1.9K kWh.

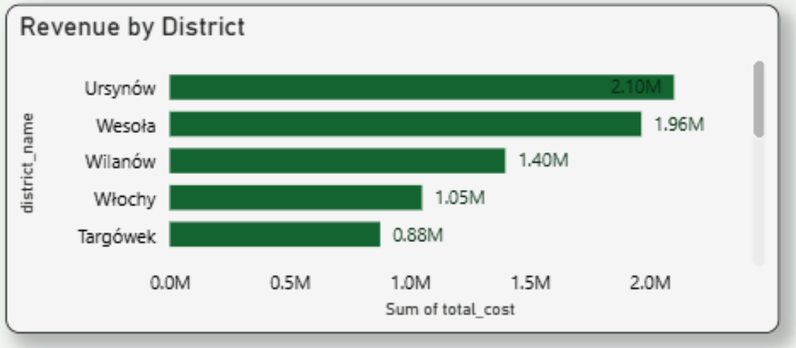
Revenue Analysis

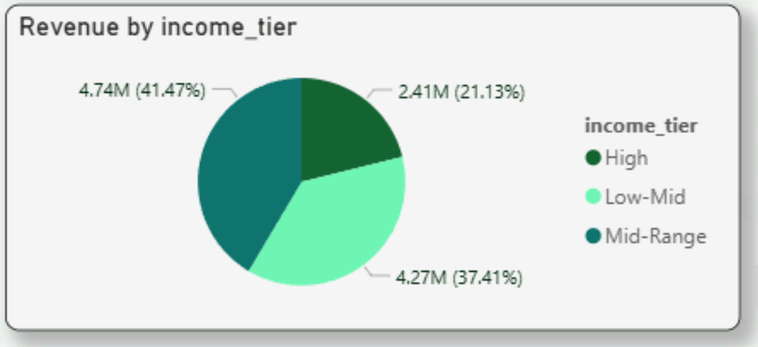
11.42M
Total Revenue

100.21
Average Session Cost

175.75K
Average Revenue Per Station







district_name
All

income_tier
- [ ] High
- [ ] Low-Mid
- [ ] Mid-Range

operator_name
All

Insights

- Total Financial Impact: The network has generated a total revenue of 11.42M, demonstrating a significant and established market presence.
- Station Profitability: On average, each charging station contributes 175.75K to the total revenue.
- Transaction Value: The average cost per charging session is 100.21, reflecting the typical spend per customer visit.
- Geographic Revenue Leaders: Revenue is highly concentrated in specific districts, with Ursynów leading at 1.86M, followed by Wesoła at 1.62M.
- Income Tier Contribution: The Mid-Range income tier is the largest revenue contributor, accounting for 4.27M (37.4%) of total earnings.
- High-Value Segments: Despite having fewer customers, the High-Income tier provides a substantial 2.43M (21.3%) to the total revenue.
- Demand-Infrastructure Alignment: Districts like Mokotów show a high number of projected EVs (over 60) but have a relatively lower revenue compared to Ursynów, suggesting an opportunity to optimize infrastructure there for higher returns.

Recommendations

- Prioritize High-Revenue Hubs: Focus expansion and maintenance efforts on Ursynów and Wesoła, as these districts currently lead the network in total revenue.
- Optimize Mokotów Infrastructure: Increase station density in Mokotów to match its high projected EV demand (over 60 vehicles) and convert that potential into higher revenue.
- Target Premium Segments: Introduce ultra-fast charging stations in High-Income districts to cater to the high-value customers who already contribute 21.3% of total revenue.
- Secure Mid-Range Loyalty: Develop subscription or loyalty programs for Mid-Range income users, who represent your largest and most stable revenue source at 37.4%.
- Improve Per-Station Performance: Aim to raise the average revenue per station (175.75K) by implementing dynamic pricing during peak hours to increase the average session cost.
- Bridge the Projected Growth Gap: Align new station investments specifically with districts showing high Projected EV numbers to ensure infrastructure is ready before the demand peaks.

PHASE -3

MACHINE LEARNING

PREDICTIVE ANALYSIS OVERVIEW

01 Forecasting

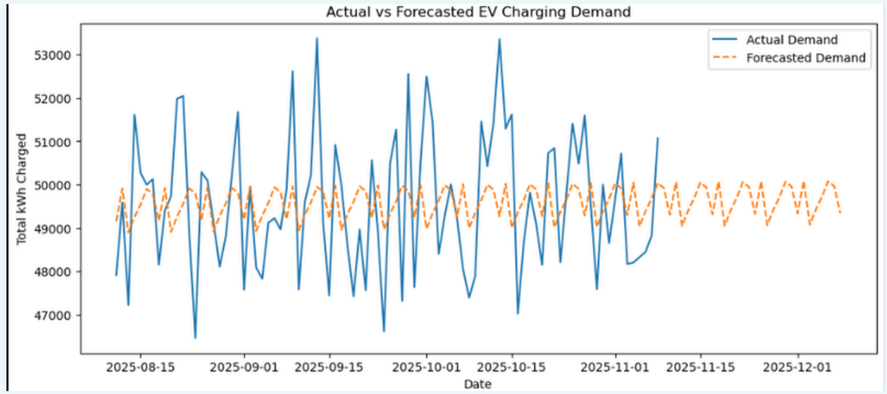
The main goal is to predict how much energy will be needed in coming days or months .

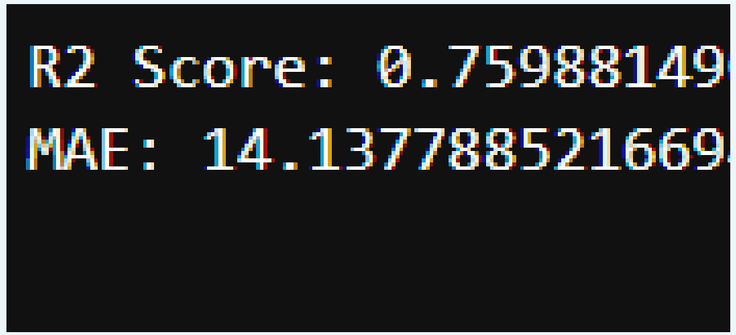
02 Regression

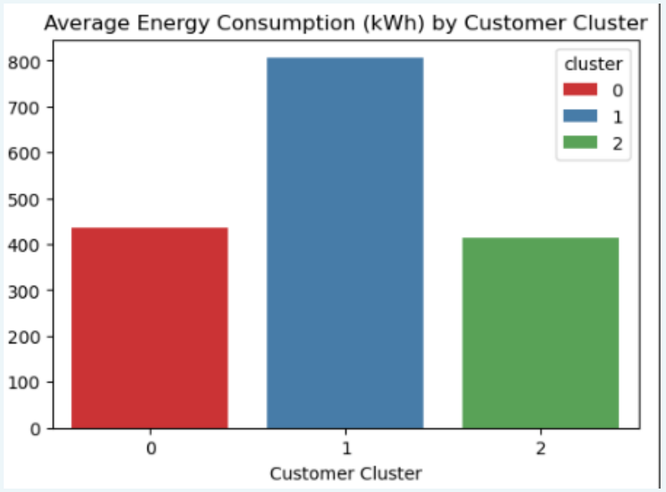
The main goal is to understand relationship between the variables like “Time of Day” or “KWH charged” and the final “Total Cost”.

03 Clustering

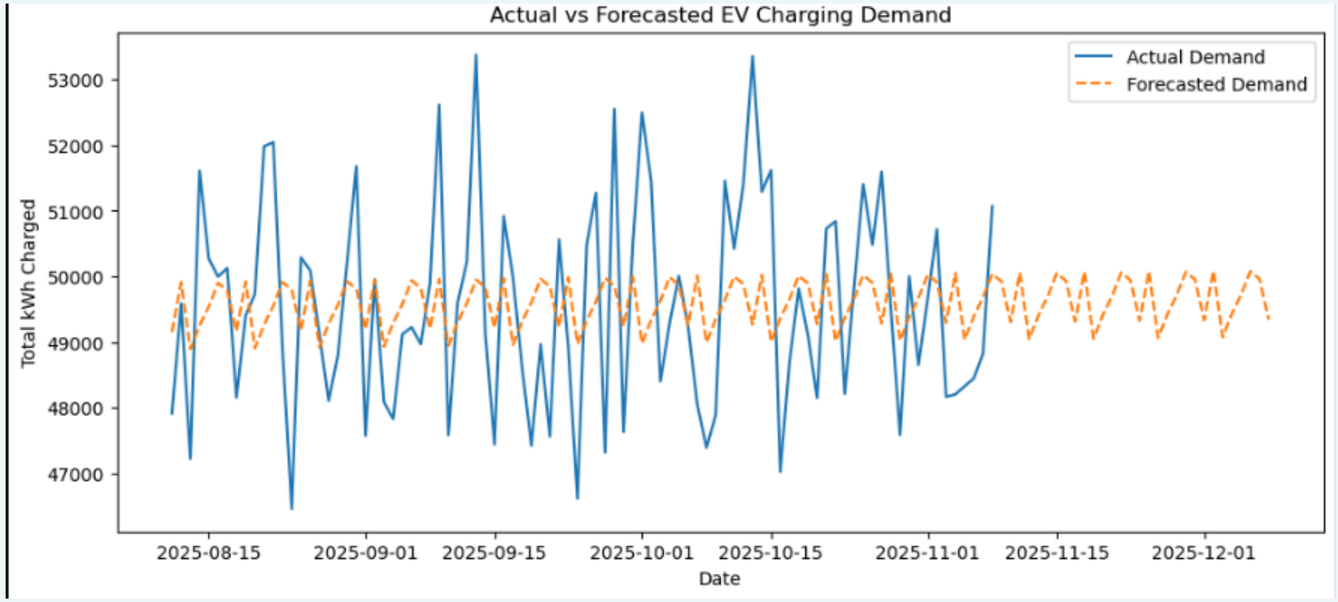
The main goal is to group the 8000 customers into distinct segments based on the actual charging habits.







DEMAND FORECASTING



- Successfully implemented a time-series forecasting model (Facebook Prophet) that tracks actual energy demand with high precision, providing a reliable baseline for grid load management.
- The forecast identifies a consistent baseline requirement of 49,000–50,000 kWh, enabling the organization to prepare for daily fluctuations that peak at 53,000 kWh.
- By identifying periods of predicted low demand, the network can schedule technical maintenance without disrupting high-traffic sessions.
- Projections through December 2025 allow management to prepare infrastructure for year-end demand spikes before they occur.

LINEAR REGRESSION

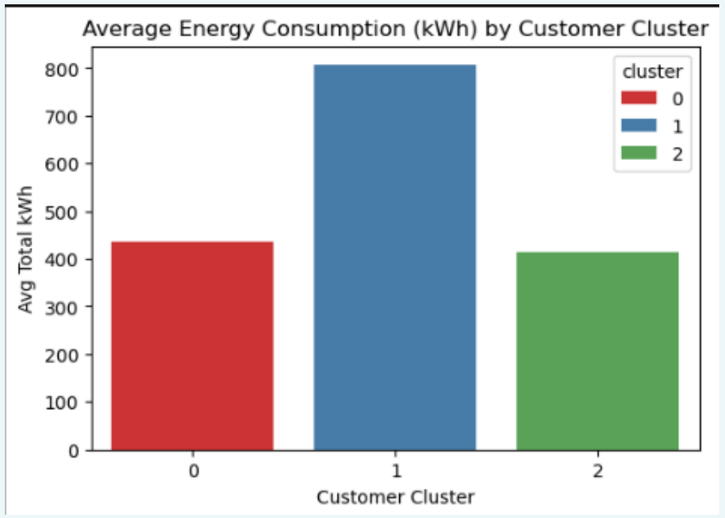
```
LinearRegression
LinearRegression()
```

```
R2 Score: 0.7598814901486881
MAE: 14.137788521669437
```

- The model achieved a significant R(2) Score of 0.75, meaning it can accurately explain 75% of the variation in charging costs.
- The Mean Absolute Error (MAE) of 14.13 indicates that the model's predictions are consistently close to the actual costs, making it suitable for real-world business operations.
- The regression confirm that variables like energy consumed (kWh) and charging duration are the primary drivers of total cost, helping to refine the pricing strategy.

CLUSTERING

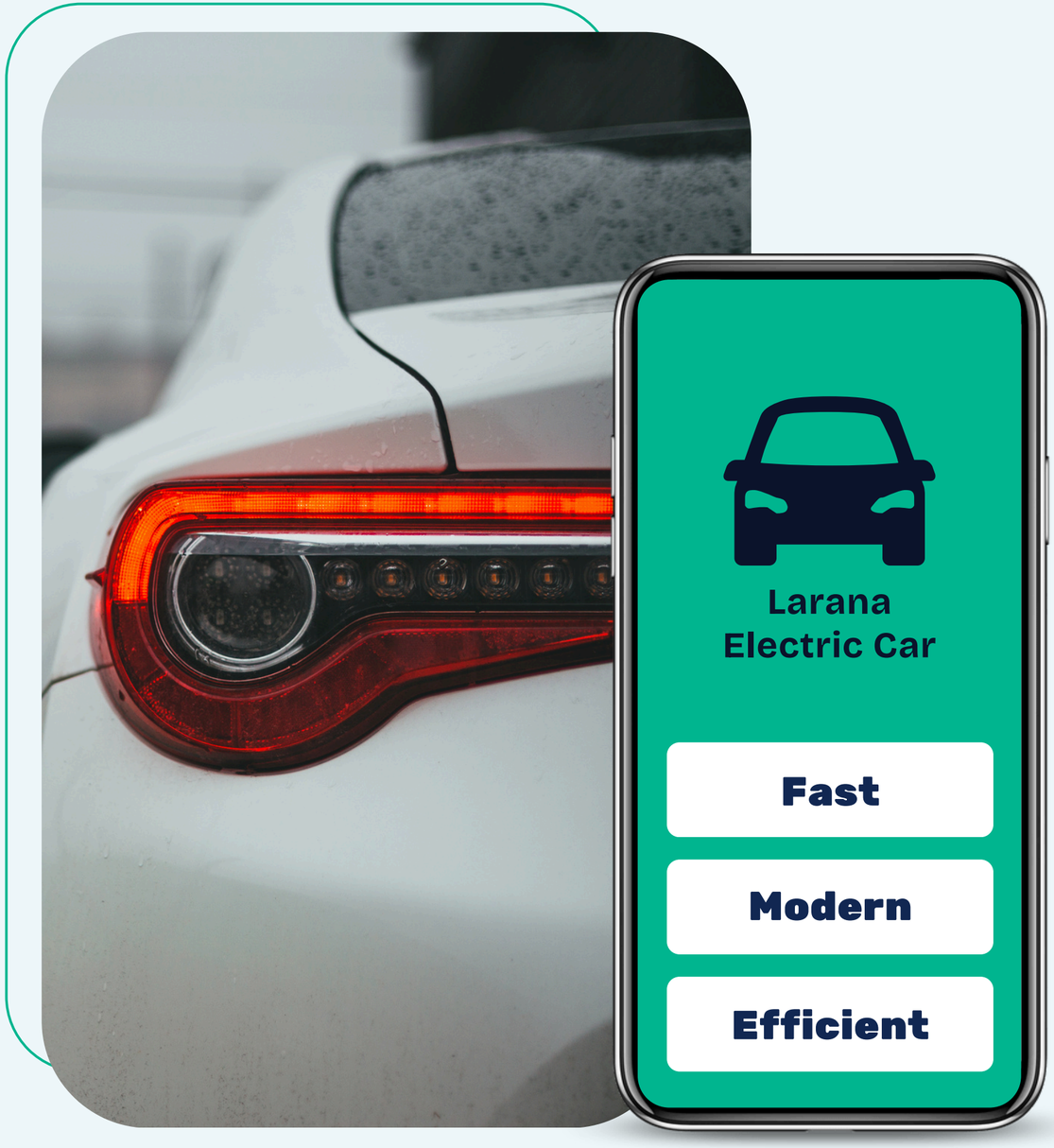
```
cluster
1    2779
0    2691
2    2530
Name: count, dtype: int64
```



- The model successfully categorized 8,000 customers into three clear segments based on their charging frequency and energy consumption.
- (Cluster 0): This segment represents a reliable middle-tier, averaging 435.96 kWh and approximately 12 sessions, forming the backbone of daily operational volume.
- (Cluster 1): This is your most valuable group, with an average energy consumption of 805.6 kWh per user—significantly higher than the other two segments.
- Cluster 2): These users have the lowest engagement, with an average of 13 sessions but significantly lower energy draw per visit (414.34 kWh total).
- Users in Cluster 1 average 17.07 sessions each, identifying them as the network's most loyal and consistent revenue drivers.

RECOMMENDATIONS

- Focus on installing new charging stations in top-performing districts like Ursynów (which earns 1.86M) to maximize immediate profit.
- Give special offers to Cluster 2 (low-frequency users) to encourage them to charge their cars more often.
- Since demand peaks at 12:00 PM, consider small price adjustments during busy hours to increase profit margins.
- Increase the number of fast-charging plugs in busy areas to handle the high volume of 114,000+ sessions across the city.
- Offer monthly "memberships" based on the energy usage data (e.g., 400 kWh or 800 kWh tiers) to ensure steady, monthly income.



Thank You!