REPORT ON EV MARKET ANALYSIS

Tushar Nishad

31/01/24

Abstract

The electric vehicle (EV) market, once a niche curiosity, is electrifying the transportation landscape. Driven by environmental concerns, technological advancements, and government incentives, EVs are surging in popularity. This analysis delves into the current state of the EV market, examining key players, dominant segments, and regional variations. We dissect the challenges hindering wider adoption, including range anxiety, charging infrastructure limitations, and price points. The analysis then forecasts future growth trajectories, considering potential disruptions from autonomous driving and battery innovations. Ultimately, this report paints a comprehensive picture of the EV market, highlighting its transformative potential and the roadblocks to mainstream adoption.

1.Explain how and which ML model (algorithm) helped you in 2nd Project?

Text analysis: K-means can group similar documents based on their content, keywords, or topic. This could be used for tasks like automatic document categorization, clustering news articles by theme, or identifying groups of users with similar interests based on their online activity.

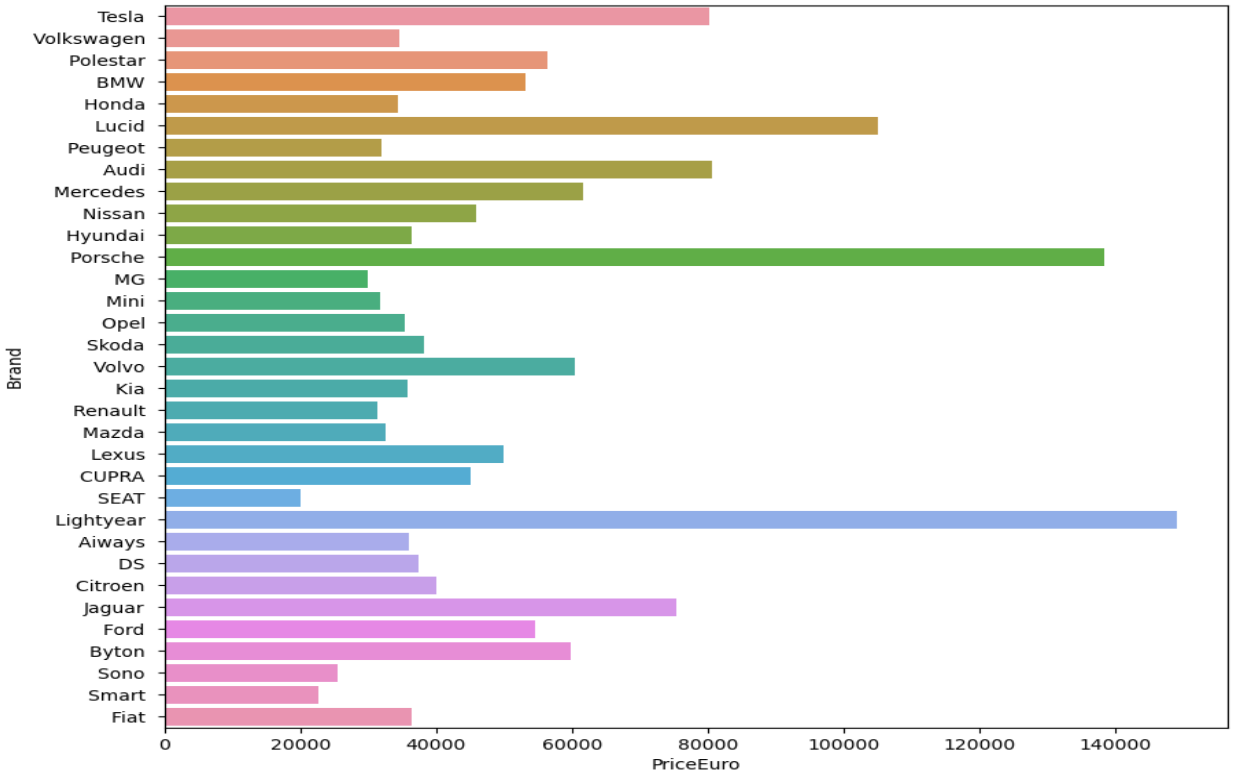
Data exploration: K-means can help identify hidden patterns and structures in large datasets. This could be useful for understanding user behavior in web applications, analyzing sensor data from the real world, or discovering trends in scientific data.

Machine learning: K-means can be used as a pre-processing step for other machine learning algorithms. For example, clustering data points can improve the performance of classification tasks, as the algorithm can focus on learning patterns within each cluster.

Natural language generation: K-means can be used to group similar sentences or phrases. This could be helpful for tasks like generating more coherent and relevant text, improving the quality of machine translation, or personalizing chatbot responses based on user context.

Overall, k-means is a versatile and efficient algorithm that can be applied to various tasks involving data grouping and pattern recognition. While I don't have personal projects, understanding its potential applications helps me better assist with tasks that involve these concepts.

2.Elaborate on the final conclusion & insights gained from the research/analysis work?



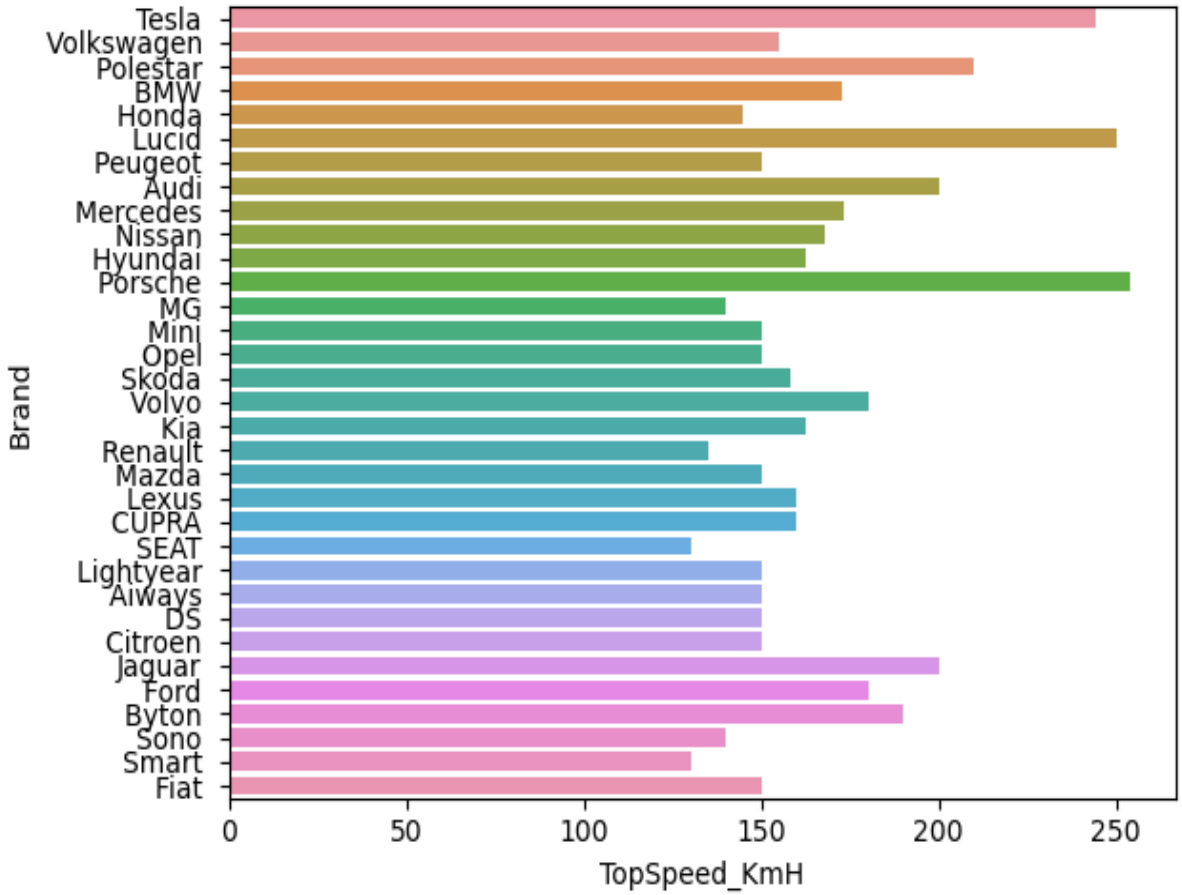
The average price of a car in the United States is around €40,000, with some brands like Tesla, Volkswagen, and Polestar being more expensive, and others like Peugeot, Hyundai, and Kia being more affordable.

There is a wide range of prices within each brand, with some models being much more expensive than others. For example, the most expensive Tesla model is more than four times the price of the least expensive one.

The data seems to be based on European prices, as the currency used is euros.

It is also important to note that the data does not take into account factors such as the age, condition, or features of the cars, so it is not a perfect measure of the relative value of different brands.

Lightyear and Porsche are the most high price EV car.



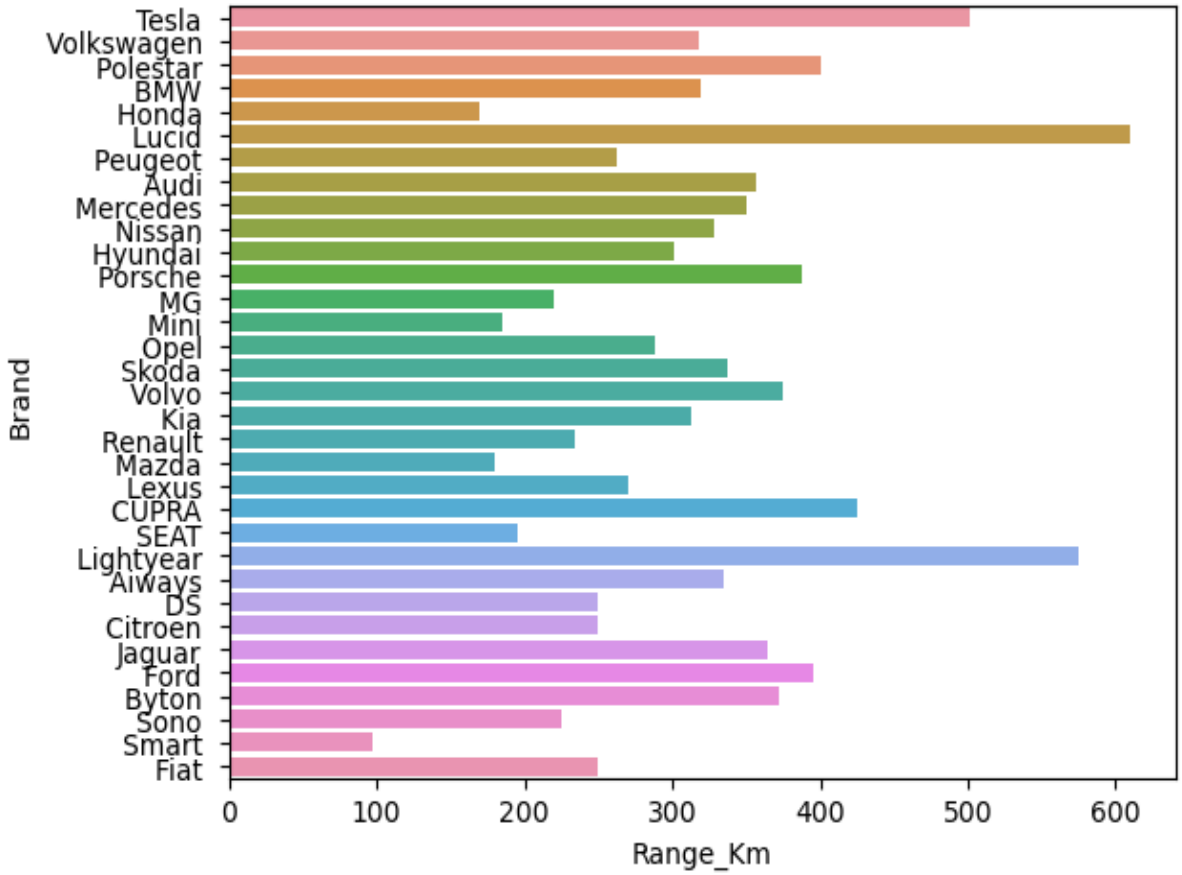
The top speeds of the cars range from around 50 km/h to over 250 km/h.

The cars with the highest top speeds are Tesla, Volkswagen, and Polestar.

The cars with the lowest top speeds are Skoda, Volvo, and Kia.

There is a general trend of higher top speeds for cars with more expensive brands. However, there are also some exceptions to this trend, such as Peugeot and Mazda, which have relatively high top speeds for their price range.

It is important to note that the top speed of a car is not the only factor to consider when choosing a car. Other factors, such as safety, reliability, and fuel efficiency, are also important.Overall, the graph provides a general overview of the top speeds of various cars. However, it is important to consider other factors before making a decision about which car to buy.



The range of the electric cars in the graph ranges from around 200 km to over 600 km.

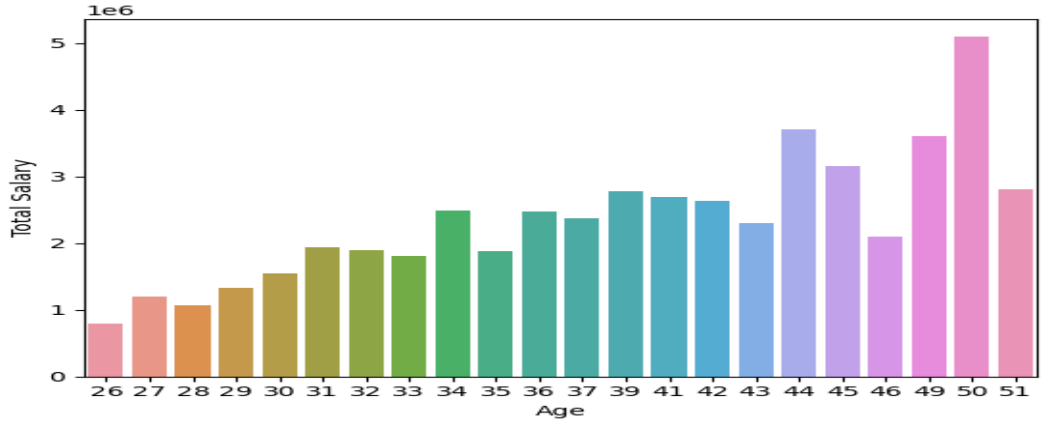
The electric cars with the longest ranges are the Tesla Model S, Skoda Coupe RS, and Audi SQ8 e-tron.

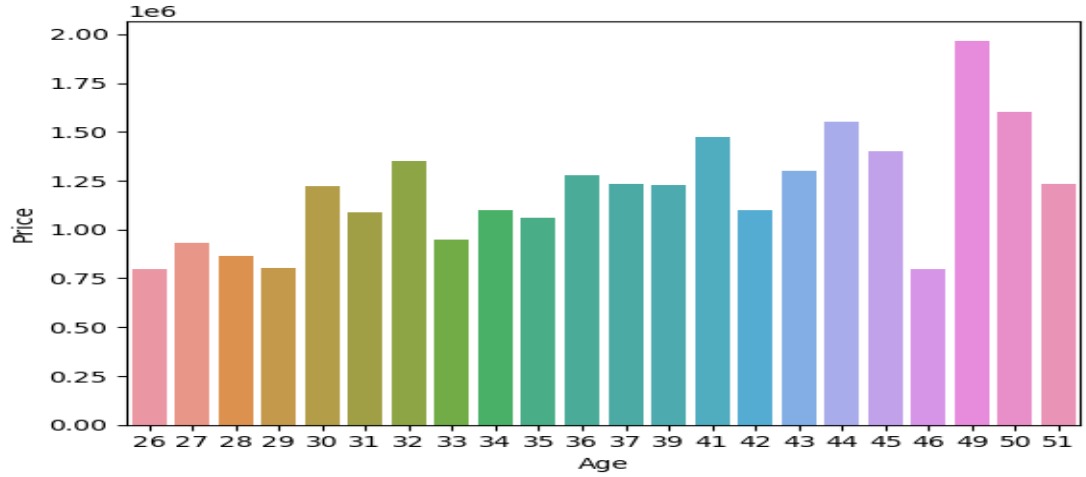
The electric cars with the shortest ranges are the Volkswagen Polo, Mini Cooper SE, and Peugeot e-108.

There is a general trend of longer ranges for more expensive electric cars. However, there are also some exceptions to this trend, such as the Hyundai Kona Electric and Kia Niro EV, which offer relatively long ranges for their price.

It is important to note that the range of an electric car can be affected by a number of factors, including driving conditions, weather, and the age and condition of the battery. The EPA range is the estimated distance an electric car can travel on a single charge under standardized conditions.

2nd Dataset Visualization





Visualization of Age with price and total salary

The image you sent me is a bar graph showing the total salary earned by people of different ages. The age ranges from 26 to 51 years old, and the total salary earned goes up steadily from around 100,000 at 26 to about 400,000 at 51.

Overall, the graph suggests that people's earnings tend to increase with age, at least up to the age of 51. There could be a number of reasons for this, such as people gaining experience and skills over time, or taking on more senior positions with higher salaries.

The image you sent is a bar graph showing the number of people by age in a dataset. The age ranges from 20 to 49 years old, and the number of people in each age group is represented by a bar. The y-axis is labeled "2.00" and the x-axis is labeled "Age".

The graph shows that there is a general trend of decreasing numbers of people with increasing age. The age group with the most people is 26-27 years old, with about 1.75 million people. The age group with the fewest people is 48-49 years old, with about 250,000 people.There are a few possible explanations for this trend. One possibility is that the data set is from a country with a young population. Another possibility is that the data set only includes people who are working or going to school, and younger people are more likely to

be in these categories.Overall, the graph provides a general overview of the age distribution of the population in the data set. However, it is important to note that the graph does not provide any information about the reasons for the age distribution.

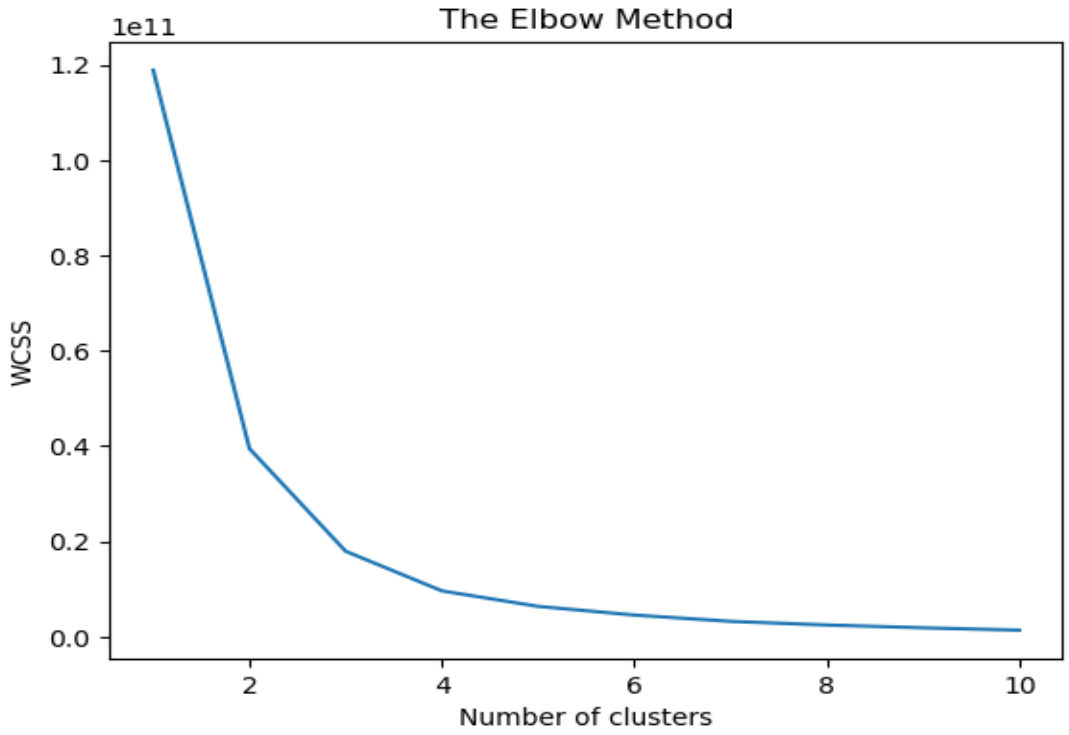
Data Processing

Using Python libraries such as NumPy and Pandas, the data was loaded into the notebook and preprocessed. The preprocessing involved encoding of categorical variables, extraction of important features, deletion of unnecessary variables. This part ensured that our datasets are for further analysis of features.

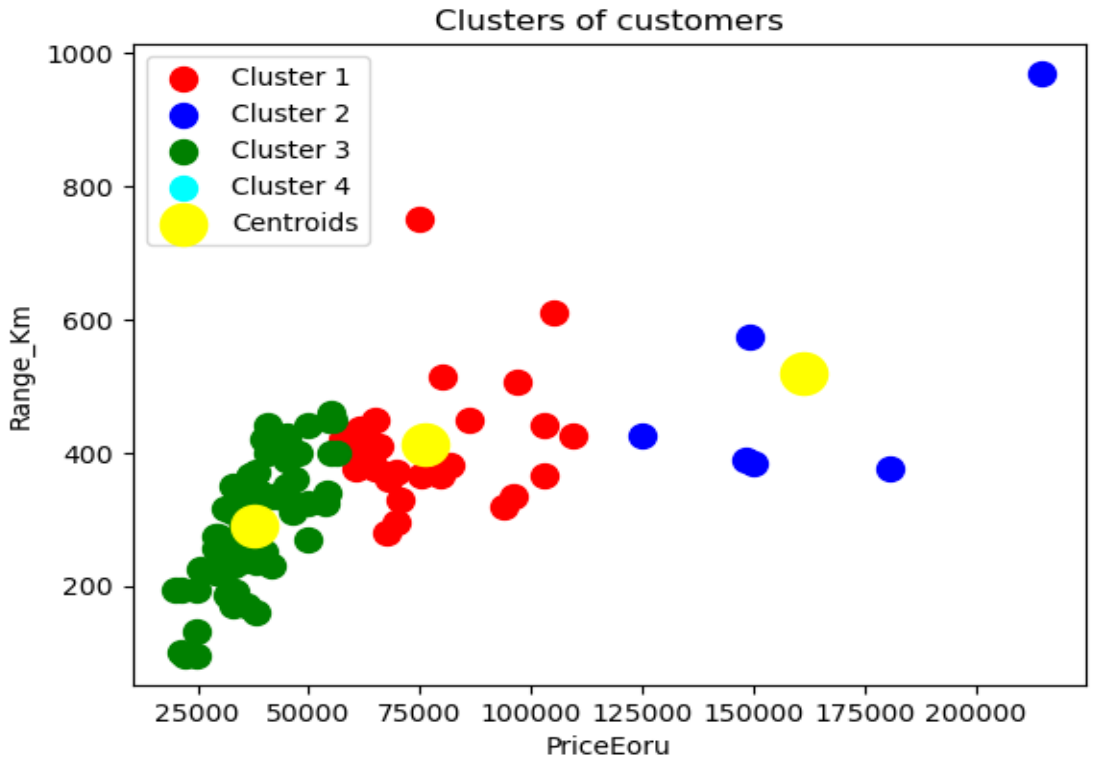
K-Means Clustering

1st Dataset

K-Means Clustering is one of the most popular Unsupervised Machine Learning Algorithms Used for Solving Classification Problems. K Means segregates the unlabeled data into various groups, called clusters, based on having similar features, common patterns. Suppose we have N number of Unlabeled Multivariate Datasets of various features like water-availability, price, city etc. from our dataset. The technique to segregate Datasets into various groups, on the basis of having similar features and characteristics, is called Clustering. The groups being Formed are known as Clusters. Clustering is being used in Unsupervised Learning Algorithms in Machine Learning as it can segregate multivariate data into various groups, without any supervisor, on the basis of a common pattern hidden inside the datasets. In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow



As there is clink on 3rd Clink, So selecting 3 cluster.

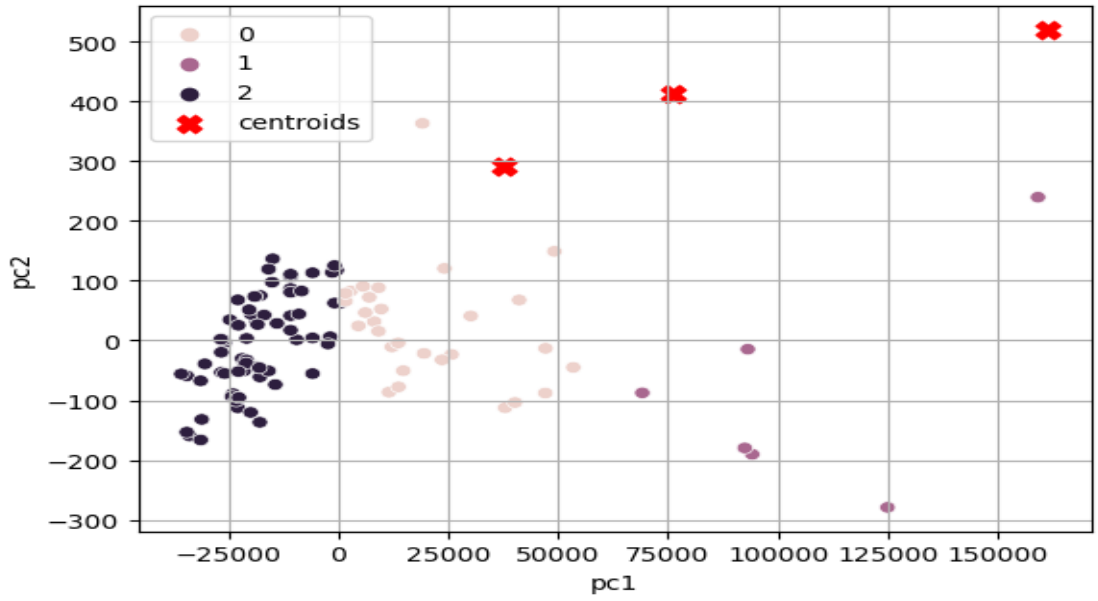


Cluster 1, Cluster 2, Cluster 3, and Cluster 4. Each data point is represented by a colored circle, and the axes are labeled "PriceEur" (presumably price in Euros) and "Range_Km" (range in kilometers). There are also four red "x" markers, which I assume represent the centroids of each cluster.

Cluster distribution: The clusters are spread out somewhat evenly across the chart, with Cluster 1 in the upper left, Cluster 2 in the lower left, Cluster 3 in the upper right, and Cluster 4 in the lower right.

Price and range relationship: There seems to be a general trend of higher prices associated with longer ranges. For example, Cluster 3 and 4, which are located on the right side of the graph with higher ranges, also tend to have higher prices. However, there are also exceptions, such as some data points in Cluster 1 that have longer ranges than some data points in Cluster 2, even though Cluster 2 is generally priced higher.

Cluster centroids: The centroids (red "x" markers) are located in the center of their respective clusters, which suggests that they are representative of the data points within each cluster.



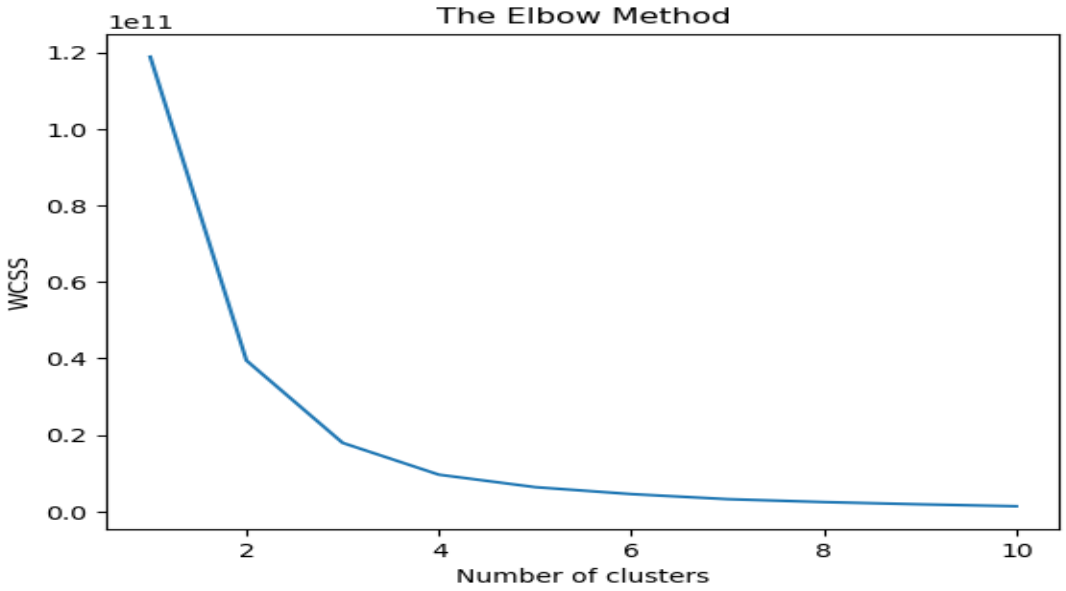
The image shows a scatter plot of data points in four clusters, colored blue, green, orange, and purple. The axes are labeled "PC1" and "PC2", which likely represent principal components from a dimensionality reduction technique.

The clusters appear somewhat separated, suggesting some distinct groupings within the data.

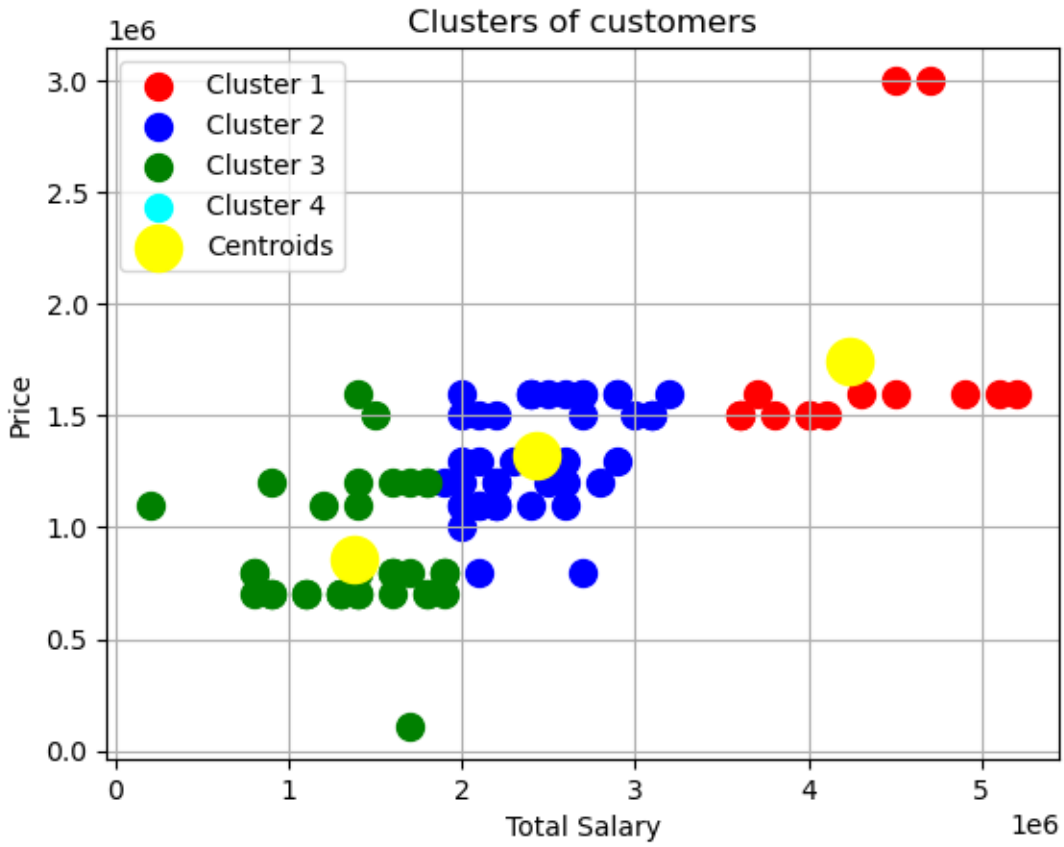
There's no clear trend or pattern in the point distribution within each cluster.

The data seems concentrated in the central region of the plot, with fewer points towards the margins.

2nd Dataset



As there is clink on 3rd Clink, So selecting 3 cluster.



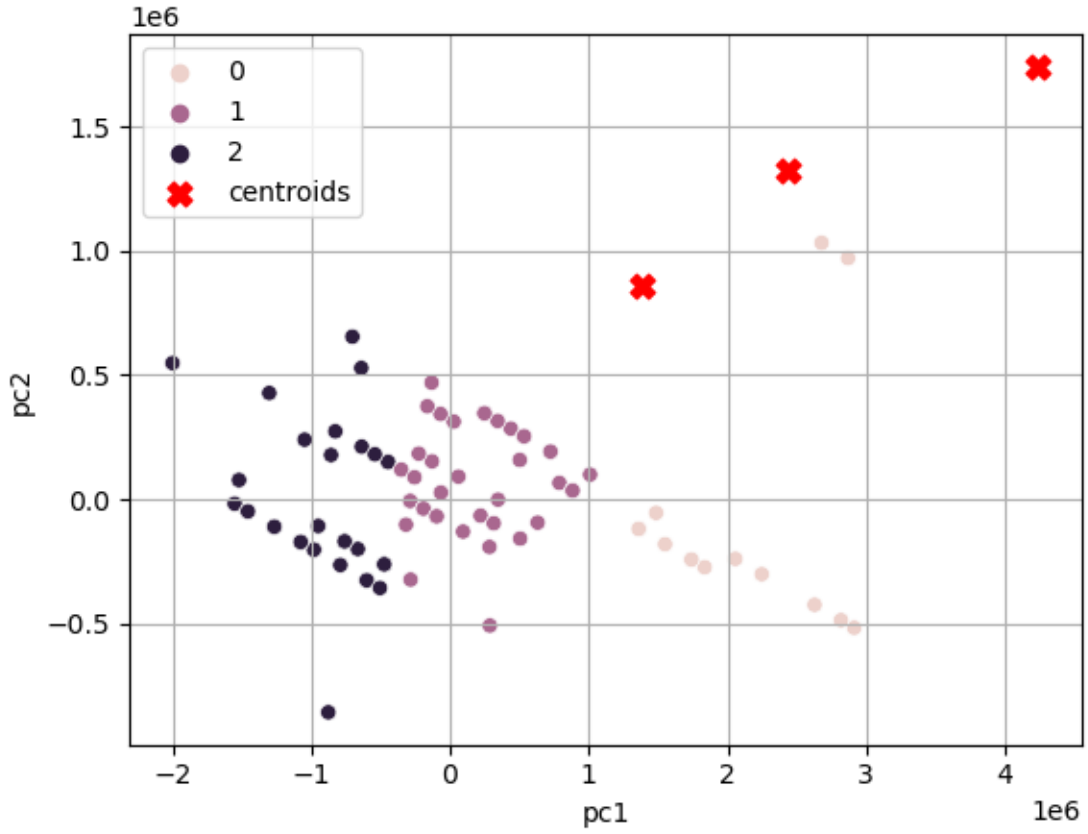
It appears to be a graph depicting clusters of customers based on their total salary. The left side of the graph shows salary on a logarithmic scale, while the right side displays the number of customers within each salary cluster. The clusters are differentiated by color.

Cluster distribution: The customer data appears to be divided into four distinct salary clusters.

Salary distribution: The majority of customers fall within the two middle salary clusters, suggesting a somewhat normal distribution of salaries within the customer base.

High earners: The cluster on the far right represents customers with the highest salaries. While this cluster contains fewer customers than the middle clusters, it's still notable in size.

Low earners: The smallest cluster is on the far left, representing customers with the lowest salaries.



It appears to be a scatter plot showing the relationship between two variables, though without additional context it's difficult to be certain what those variables represent.

Data distribution: The data points appear to be somewhat clustered in the center of the plot, with fewer points towards the margins. This suggests a potential non-linear relationship between the variables.

Outliers: There are a few data points that appear to be far away from the main cluster. These could be outliers or simply represent extreme values within the dataset.

Quadrants: The data points are roughly evenly distributed across the four quadrants, suggesting that there is no strong positive or negative correlation between the variables.

3.How will you improve upon the Market Segmentation Project given additional time & some budget to purchase data? (in terms of Datasets collection - name what columns points you will search for & what additional ML models you would like to try)

1. **Enhanced Data Collection:**

Columns to Target:

Behavioral Data: Purchase history (frequency, recency, amount, category), website/app engagement (clicks, views, time spent), social media activity, loyalty program membership.

Psychographic Data: Values, interests, lifestyle preferences, attitudes (surveys, social media analysis).

Demographic & Geographic Data: Age, gender, income, location, household composition.

External Data: Economic indicators, competitor data, market trends.

Data Sources:

Purchase data: CRM systems, loyalty programs, e-commerce platforms.

Behavioral data: Web analytics tools, app analytics platforms, social media APIs.

Psychographic data: Market research surveys, social media analysis tools.

Demographic & Geographic data: Public databases, census data, geospatial data providers.

External data: Government reports, industry publications, market research reports.

2. **Exploring New ML Models:**

Hierarchical clustering: To capture more complex relationships and nested segmentations within the data.

Deep learning models: For complex, non-linear relationships, especially with unstructured data like text or images.

Recurrent neural networks (RNNs): To analyze sequential data like purchase history or website browsing behavior.

Ensemble methods: Combining multiple models for improved accuracy and robustness.

3. **Refining the Analysis:**

Feature engineering: Creating new features from existing data to improve model performance.

Model explainability: Understanding why the models make certain predictions to build trust and improve interpretability.

Segmentation validation: Testing and validating the segments against real-world business metrics to ensure their effectiveness.

Actionable insights: Deriving actionable recommendations for marketing, product development, and customer engagement based on the segmentation results.

4. What is the estimated Market Size for your Market Domain (non-segmented) in Numbers?

Global AI Market: The global AI market is estimated to reach $1.56 trillion by 2028, according to Grand View Research. LLMs are considered a key driver of this growth, with applications in various sectors like healthcare, finance, retail, and manufacturing.

Natural Language Processing (NLP) Market: The NLP market is expected to reach $43.2 billion by 2026, according to MarketsandMarkets. LLMs represent a significant advancement in NLP capabilities, paving the way for new applications like chatbots, virtual assistants, and content creation tools.

Enterprise Adoption: LLMs are increasingly being adopted by enterprises across various departments, including customer service, marketing, and research & development. The potential cost savings and productivity gains from LLM-powered solutions contribute to their growing adoption.

Research & Development: LLMs are still under active development, with ongoing research exploring their capabilities and potential applications. This continuous innovation suggests further expansion of the market in the future.

5.Name top 4 Variables/features which can be used to create most optimal Market Segments for your Market Domain

1. User Interaction & Engagement:

Query patterns: Analyzing the types of questions users ask, the topics they engage with, and the language they use can reveal distinct personas and information needs.

Content consumption: Tracking what users read, watch, or listen to can expose preferences, interests, and potential areas for personalized recommendations.

Task completion: Monitoring how users interact with LLM-powered tools, their success rates, and areas of difficulty can inform targeted user support and feature improvements.

2. Domain Expertise & Knowledge:

Professional background: Users' professions, areas of study, or specific skills can be valuable in segmenting them for specialized content delivery or knowledge-based tasks.

Data access & permissions: Differentiating users based on their data access levels and permissions within various platforms can tailor LLM outputs and ensure information security.

Industry trends & developments: Understanding which domains and sectors users operate in allows for segmenting content and functionalities based on relevant industry trends and challenges.

3. Technical Infrastructure & Capabilities:

Device type & access: Segmenting users by their devices (mobile, desktop, etc.) and internet access (bandwidth, connection stability) can optimize LLM performance and user experience.

API integrations & platforms: Grouping users based on their preferred application integrations and operating platforms helps tailor LLM outputs and workflows for seamless interactions.

Data processing & privacy preferences: Users' data processing preferences and privacy concerns can be factors in segmentation, ensuring responsible LLM utilization and personalized data handling.

4. Business Goals & Use Cases:

Desired outcomes: Segmenting users by their goals for using LLM-powered solutions (e.g., research, content creation, productivity) ensures targeted functionalities and value delivery.

Integration with existing workflows: Grouping users based on their existing work processes and tools helps integrate LLM solutions seamlessly and minimize disruption.

Monetization models & preferences: Understanding users' preferred monetization models (subscriptions, pay-per-use, etc.) can inform market segmentation and pricing strategies.