**AC-5002 R Programming Assignment**

**Mehul Kulkarni - 2543907**

**PART 1:** Most popular namesLoad the babynames tibble. Use dplyr to find out which were the three most popular names given tobabies of male sex during the 20th century (1900-1999) in the USA. Use ggplot2 to produce a singlegraph that shows how the popularity these three names changed over the course of the century.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

install.packages(c("tidyverse", "babynames", "dplyr", "ggplot2")) #This installs all the necessary packages

library(tidyverse)

library(babynames)

library(dplyr)

# Filter for most popular names between 1900 and 1999

most\_popular\_names <- babynames %>% #here i have created a new table named most\_popular\_names that will extract data from babynames

filter(sex == "M", between(year, 1900, 1999)) #filter function allows to filter the dataset, here i have filtered sex, and year columns

View (most\_popular\_names) #View shows the most\_popular\_names table

# Find the three most popular names

three\_popular\_names <- most\_popular\_names %>% #Created three\_popular\_names table

group\_by(name) %>% #The group\_by function is used to group rows of a data frame based on one or more variables

summarize(total = sum(n)) %>%

arrange(desc(total)) %>%

head(3)

View(three\_popular\_names)

# Filter data for the three most popular names

filtered\_data <- babynames %>%

filter(sex == "M", name %in% three\_popular\_names$name, between(year, 1900, 1999))

#This section filters the original babynames data frame to include only rows where 'sex' is "M," the 'name' is one of the three popular names, and the 'year' is between 1900 and 1999.

# Create and customize the ggplot

myplot <- ggplot(filtered\_data, aes(x = year, y = n, color = name)) +

geom\_line() +

labs(

title = "Three most popular baby names in 20th century",

x = "Year",

y = "No. of occurrences",

color = "Name"

) +

theme\_minimal()

#This section uses the ggplot function to create a line plot (geom\_line()) based on the filtered\_data. The aesthetics (aes) are set to use 'year' on the x-axis, 'n' (number of occurrences) on the y-axis, and color the lines by the 'name' variable

# Display the plot

print(myplot)

#Finally, the print function is used to display the plot created in the previous step. The resulting plot shows the trends of the three most popular baby names for males in the 20th century

A graph on a computer screen

Description automatically generated

The above line graph shows the three most popular names (James, John and Robert) and tells how many times these names have occurred in the year 1900-1999. In the graph we have chosen Year on X axis which is between 1900-1999, Y axis we have put No. of occurrences which means total number of times these name have been chosen in the following years and the different colors on line distinguishes different names of babies, where Red represents James, blue represents Robert and green represents John.

**Part 2: Data Exploration.**

In this assessment I have chosen a dataset which tells about EVehicle Specification and Prices. a comprehensive collection of data scraped from EV Database which I have downloaded from Kaggle.com. This dataset presents a detailed exploration of various electric vehicles, providing valuable insights into their specifications, pricing, and performance metrics.

This dataset consist of 360 observations and 9 columns from which 6 of them are integer, 1 string, 1 decimal and 1 consists of links of the car. This dataset includes a wide range of data that is essential for comprehending the state of electric mobility. Every entry has information about the battery capacity, model name, efficiency ratings, fast-charging capabilities, cost in Germany, range, top speed, and acceleration from 0 to 100 km/h. There is also a link to the EV Database for more details.

Columns in the dataset:

* Battery (integer data type): The capacity of the vehicle's battery in kilowatt-hours (kWh).
* Car\_name (string data type): The model name of the electric vehicle.
* Car\_name\_link : A direct link to the corresponding page on EV Database for more in-depth information.
* Efficiency (integer data type): The energy efficiency rating of the vehicle in watt-hours per kilometer (Wh/km).
* Fast\_charge (integer data type): The fast-charging capability of the vehicle in minutes for a certain charging percentage.
* Price.DE.(integer data type): The price of the electric vehicle in Germany.
* Range(integer data type): The driving range of the vehicle on a single charge in kilometers.
* Top\_speed (integer data type): The maximum speed the vehicle can achieve in kilometers per hour.
* Acceleration..0.100(decimal data type): The acceleration time from 0 to 100 kilometers per hour.

Using the following columns I have derived various visual and descriptive information’s about cars compared each of the features and got a best suitable vehicle for every requirement type. Generally, there are models released in 2021 and later and there are also models that can be ordered in 2024 which is also the reason for some of the rows mentioning ‘NA’ for some columns. For a more detailed review, I recommend you to visit the website where the data is available. Therefore I have tried wrangling the dataset by using dplyr which we will see in upcoming analysis.

To give the best report about the data I have used various packages in R studio that are tidyverse, dplyr, ggplot, readr and ggrepel.

* Tidyverse: It is a collection of R packages that share a common principle, it provides best tools for data analysis making it easier to manipulate, visualize and model the data.
* Ggplot: This package is for creating static, animated and interactive data visualizations, users can plot complex graphs
* Dplyr: This package has a set of functions for data manipulations which includes selecting, arranging, summarizing, filtering and joining data frames.
* Readr: This package is used to read and write rectangular data in plain text files.
* Tibble: Tibbles are part of the tidy data principles and have some advantages over traditional data frames.

**Visual representations of various information extracted from data.**

**Top 10 Efficient cars with their respective prices :**

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

In this bar graph I have tried to show which cars have the most efficiency and what are their prices. The x axis gives us the name of the EVcars, Y axis tells us about the efficiency of each car, and bar on the right side shows the price range where the darkest shade is the lowest and the lighter shades is the highest price range.

From this we get to know that the car ‘Mercedes eVito Tourer Extra-Long 90 kWh’ is the most efficient car out of all the segment having the efficiency of 295 and Price of 65140.

**Relation of car and their top speed arranged in price low to high.**

**A screenshot of a computer

Description automatically generated**

**A graph on a black background

Description automatically generated**

The following line graph gives us a detailed information about the top 10 cars and their top speeds. We got this by keeping car\_name column on the x axis, and top speed column on the Y axis. And from the following graph we get to know that ‘Maserati GranTurismo Folgore’ is the car with highest top speed (320) and ‘  
Porsche Taycan Turbo S’ has the lowest top speed amongst the 10 best.

**Graph for the cars having prince under 45000, top speed less than equal to 200 and acceleration more than 7**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

The following graph gives us detailed information about the cars which have prince under 45000, top speed less than equal to 200 and acceleration more than 7. By reading this we get to know that out of all ‘Volkswagen ID. Buzz Pro’ has the highest acceleration as the point is the biggest, has price around 65000 and top speed not too less than 150.

**Top 10 cars with battery between 70 to 80 and Range is <=425.**

A screenshot of a computer

Description automatically generated

A graph of blue color

Description automatically generated with medium confidence

The following bar graph states different cars having battery between 70 to 80 range should be less than or equal to 425. We got this by keeping car name on x axis and range on y axis. The shade bar on the right side of the graph tells about the battery of each vehicle where the darkest blue is the lowest battery (70) and light blue is highest battery (80). From the graph we get to know that ‘Volvo C40 Recharge Single Motor ER’ has the best battery closest to 80, and range is around 420.

**Cars with worst fast charging capabilities.**

**A screenshot of a computer

Description automatically generated**

**A graph with lines and dots

Description automatically generated**

The above line graphs gives us explanation about the cars having worst fast charging capability in a car. We got this graph by keeping car name on X axis and fast charge column from dataset on Y axis. And by reading the graph we get to know that ‘Hyundai IONIQ 6 Long Range 2WD’ has the worst fast charging capability than all the other cars. It takes 1290mins to charge the car which is the longest of all.

**Long lasting vehicles which is based upon the weighted scores of battery, efficiency range and price**

A screenshot of a computer

Description automatically generated

A graph of different colored squares

Description automatically generated

The following graph categories the cars which could be said as the best pick for normal day to day use. This is categorized by taking into account of four factors, and they are battery, efficiency, range and price. I have given a certain weighted score for each and derived vehicles which would be suitable for this category. The weighted score is calculated as Battery \* battery\_weight +

Range \* range\_weight +

Efficiency \* efficiency\_weight +

Price.DE. \* price\_weight) / sum(c(battery\_weight, range\_weight, efficiency\_weight, price\_weight)). And the weighted score given to each of the components were as follows:

battery\_weight = 0.3,

range\_weight = 0.3,

efficiency\_weight = 0.2,

price\_weight = 0.2.

Hence, the best car for day to day and long lasting use is ‘Lucid Air Dream Edition P’ having highest battery of 118, range of 685, efficiency of 177 and comes in price 218000.

**Best EVSport Car category**

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

The following graph categories the cars which could be said as the best in sports. This category has taken by choosing several factors such as top speed, acceleration, fast charge and price. I have given a certain weighted score for each and derived vehicles which would be suitable for this category and the weighted scores were as follows:

speed\_weight = 0.45,

acceleration\_weight = 0.35,

charge\_weight = 0.2.

Followed by the formula :

sports\_car = (Top\_speed \*speed\_weight +

acceleration..0.100. \* acceleration\_weight +

Fast\_charge \* charge\_weight) / sum(c(speed\_weight, acceleration\_weight, charge\_weight))

After which from the graph we get to know is ‘Porsche Taycan 4S Plus’ is the best in sports EVcar out of the top 10 having top speed of 250, acceleration of 4, fast charge 1180, and price 120081.

**Conclusion:**

After a comprehensive analysis of the EVcars dataset various insights have emerged regarding the electric vehicles. We were able to derive plenty information about every aspect of vehicle which were battery, efficiency, price, range, top speed and acceleration and could also categorize each vehicle in each aspects, this could also help future buyers for best pick of their vehicle. Also giving a helping hand for brands to invent vehicles asper the customer demands. In conclusion this information and report is extremely useful in making educated decision for buying and selling of Evehicles.