**Bhakti Patil, Shashank Rewankar, Prithvi Bhatt, Arnav Laturkar**

**Project - Phase III: Dashboard Implementation**

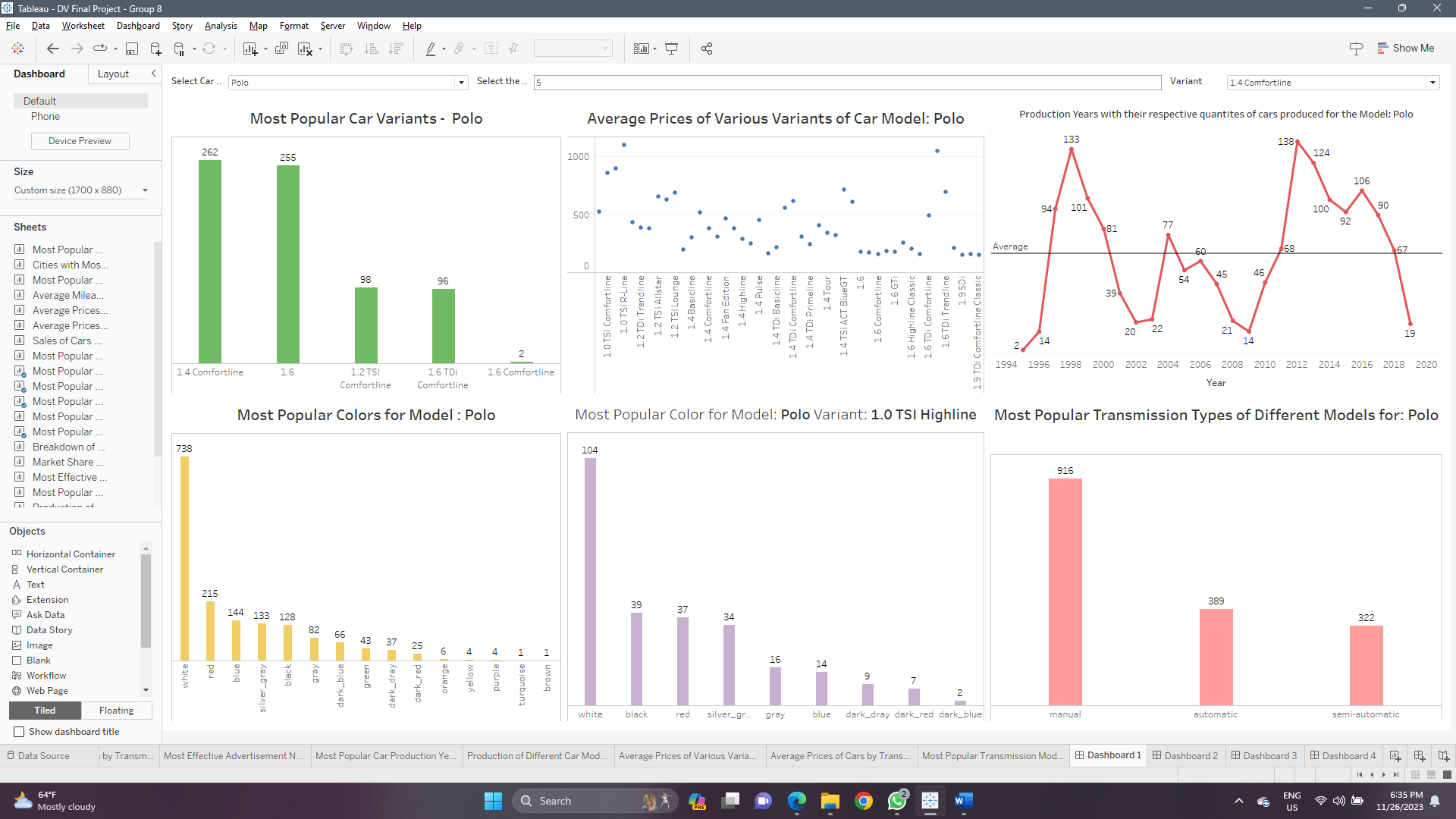
**IFT 533- Data Visualization and Reporting for IT**

**Prof. Asmaa Elbadrawy**

**November 26, 2023**

**Section 1: Dashboard**

**Dashboard 1: Visualizing information related to car models and their variants**



In the above dashboard, we are visualizing and exploring various insights into the different car models in their variants. There are 6 plots included in this dashboard which display the most popular car variants of a model and their average prices, the different production years of the selected car model, the most popular colors for the selected car model, most popular colours for a particular variant of a model, and the most popular transmission modes for the car model selected. This dashboard is interactive as it includes three different user-interactivity controls – selecting the car model, selecting the car variant and selecting the number of variants to display.

**Dashboard 2: Visualizing information related to the cities where the cars were purchased**

A screenshot of a computer

Description automatically generated

In the above dashboard, we are visualizing the information about the cars purchased in different cities. This dashboard includes three plots – the first two showing the most popular car models and transmission modes of the car in a city respectively, and the third is a bar chart that shows the cities that have the greatest number of cars purchased. This dashboard is also interactive as it has three user interactivity controls – the first one for selecting the city to show the popular car models and transmission modes used, second for limiting the number of car models to show, and the last one is used for the third plot, where a user can enter an integer input to identify cities that have greater amounts of cars purchased than the user input by coloring the bars accordingly.

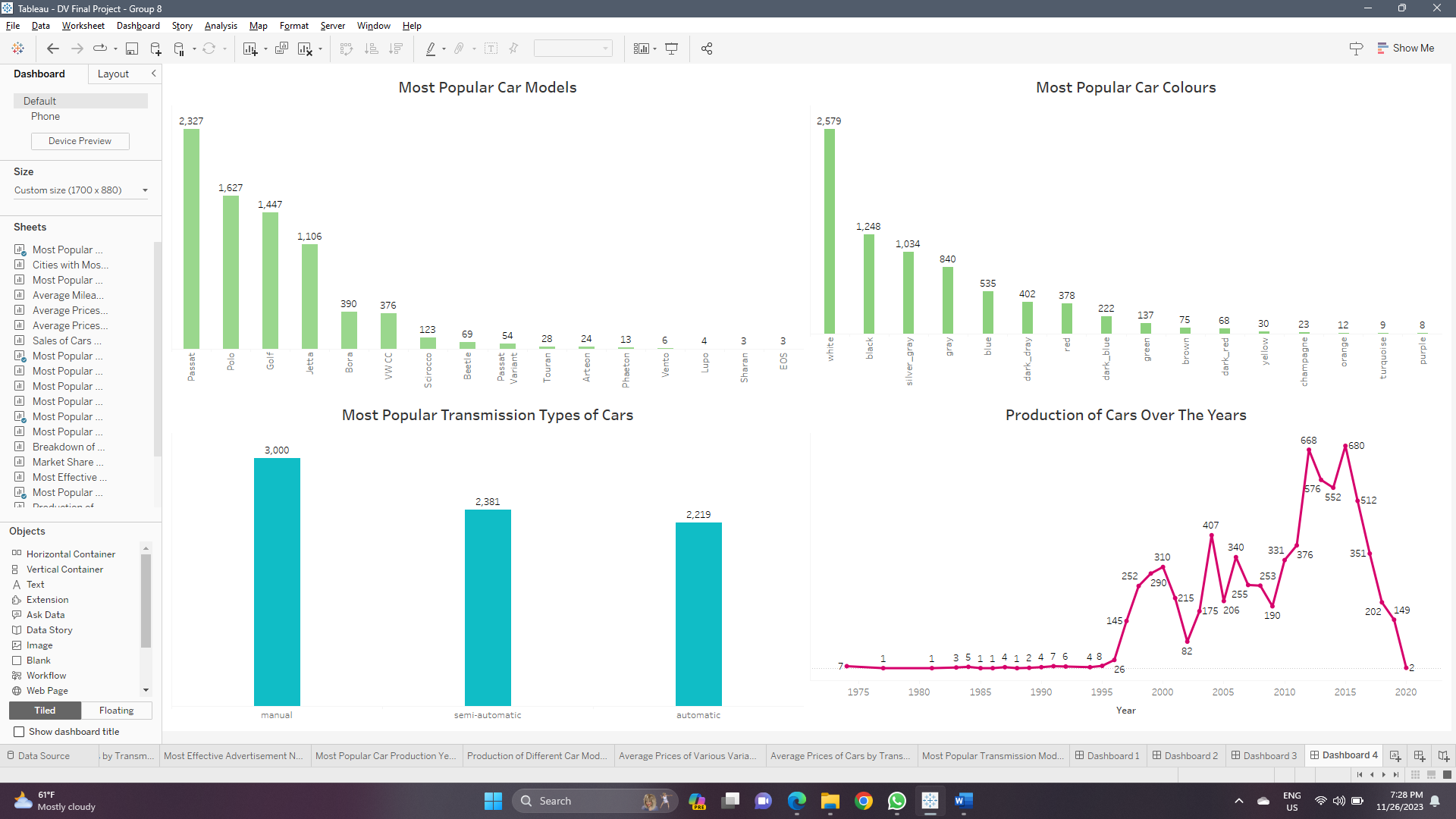
**Dashboard 3: Visualizing market shares of cars and average prices by fuel types and transmission modes**

A screenshot of a computer

Description automatically generated

This dashboard contains four different plots – first a dot plot showing average prices of cars based on fuel type used by the car and the transmission mode of the car. The second plot is a scatter plot comparing the average prices of the car and their respective mileages, then another plot is a bar chart showing the average prices of the cars based on just the transmission mode used, and lastly a pie chart showing the market shares of the cars based on the two factors – transmission mode and fuel types.

**Dashboard 4: Visualizing key metrics across the entire dataset**



In this dashboard, we are showing the key information of the cars purchased in our dataset. There are four plots in this dashboard visualizing the most popular car models, most popular car colours, most popular transmission modes of cars and lastly visualizing the different production years of all the cars that were produced.

**Section 2: Dataset**

**Online Automotive Sales Statistics'23 (Volkswagen) :**

There are a variety of publicly available datasets that provide statistics on Automotive Sales (on Kaggle and related websites), however there aren't many Volkswagen-specific datasets available online. The data for this dataset came from Turkey's leading online real estate and vehicle sales marketplace. The data solely applies to Volkswagen automobiles and covers the three months from January 1 to March 31, 2023.

This dataset contains data scraped from 7216 customers in order to examine the automotive market. A row in the dataset correlates to other columns for various car types, costs, color, and locations. Numerous null entries are present in the Date and Fuel columns and can be pre-processed before real visualizations are made. The dataset provides useful information about Volkswagen automotive sales and advertising tendencies in Turkey during the first quarter of 2023. The data can be utilized to identify patterns and trends in customer behavior, such as the most popular models, transmission styles, and fuel kinds. The data might also be used to discover which cities have the highest demand for Volkswagen vehicles and to evaluate the effectiveness of advertising campaigns.

The Online Automotive Sales Statistics'23 (Volkswagen) dataset includes 7216 records (rows) and 13 attributes (columns).

Attributes with what they mean, their Data types and Domain values:

1. Customer ID - Ordinal, Domain - Integer [0 - 7216]

A unique identifying number assigned to the person who is advertising.

1. Advertisement number - Categorical, Domain - Integer [195460036 – 777835401]

Unique identification for the AD

1. Brand - Categorical, Domain - String

Brand name of the car.

1. Model - Categorical, Domain - String

Model name of the car.

1. Variant - Categorical, Domain - String

Version of the car.

1. Year - Interval, Domain - Date type

Year when the car was developed and ready for use.

1. Kilometer - Categorical, Domain - Integer

Gives information about the distance the car has traveled.

1. Color - Categorical, Domain - String

Specifies the car color.

1. Transmission - Categorical, Domain - String [Automatic, Manual]

Takes two values and provides the transmission type of the car.

1. Fuel - Categorical, Domain - String [Diesel, Gas, Gasoline]

Specifies the type of fuel the car uses to run on.

1. City - Categorical, Domain - String

Specifies the city where the car is advertised.

1. AD Date - Interval, Domain - Date Type

Date information of the ad release.

1. Price - Ordinal, Domain - Integer

Price of the car.

**Data Pre-Processing done on this dataset:**

1. Removing null values from the dataset for the attribute - brand.

2. Aggregating and summarizing the data for the attributes - model, year, color, fuel and transmission, kilometer and city.

3. Removing null values for attributes - kilometer and city.

4. Removing suffixes in the attribute price and converting its data type to floating point value.

**Section 3: Dashboard Users**

Following is the list of prospective users for our planned dashboard:

* **Sales Managers**: They can use the dashboard to monitor the performance of sales, segment customers, predict and set sales targets, keep track of inventory levels and also understand geographical effects on the sales of cars to make decisions accordingly.
* **Marketing Managers**: They can use the dashboard to swiftly assess the overall situation of the marketing efforts, evaluate the effect of the marketing campaigns, optimize lead generation strategies and understand the amount of budget allocation requirements.
* **Operations Managers**: They can use the dashboard to monitor the various inventory levels, identify popular vehicles, anticipate demands of various vehicles, streamline the process of order fulfillment and optimize the entire supply chain process by identifying bottlenecks.
* **Business Owners**: Business Owners can monitor the key performance indicators (KPIs) such as total sales, revenue and profit margins, financial metrics such as expenses and profits, forecast sales trends, monitor customer feedback to improve upon efficiency of the overall business and make key strategic decisions to expand the business.

**Section 4: Questions**

1. Which are the most popular car models?

2. Which cities have the most customers that purchase cars? What are the most popular car models there? Which cities have the fewest customers?

3. What are the average distances traveled by cars that use different transmission modes and fuel types?

4. How does the average price of gasoline cars compare to diesel and hybrid cars?

5. How does the average price of the different car models compare to their respective mileages?

6. Which month of advertisement dates had the highest effect on the sales of cars?

7. Which car colors are the most popular among customers? What are the most popular variants of the different car models?

9. What are the most popular colors for each car model? What are the most popular colors for each variant of a car model?

11. What are the most popular transmission types in cars purchased? What are the counts of fuel types used by the car by different transmission types?

12. Which advertisement numbers brought in the highest number of car sales?

13. Which year of production of cars is the most popular among customers? What are the distributions of years of production of different car models?

15. What are the average prices of the different variants of the same car models?

16. What are the averages prices of cars based on types of transmission mode used by the car?

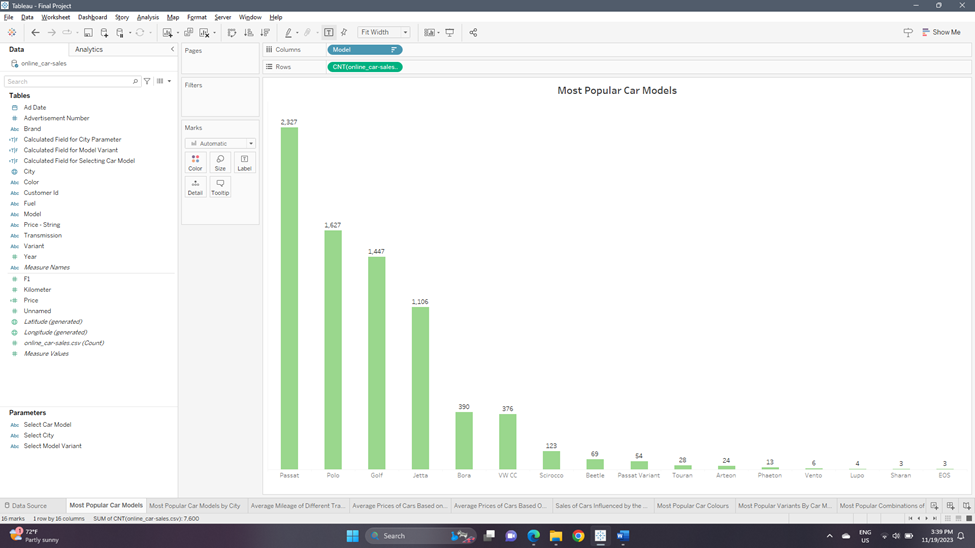
17. Which transmission modes are the most popular in each city?

**Section 5: Plots**

Plots explaining the questions which they address:

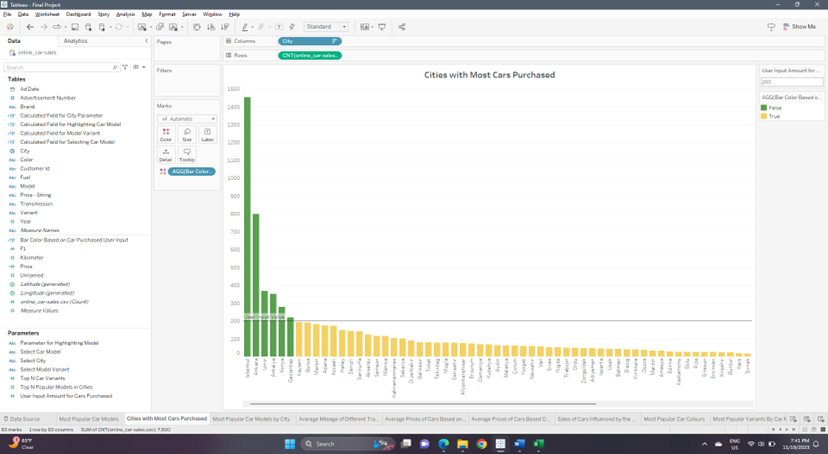
**1. Which are the most popular car models?**

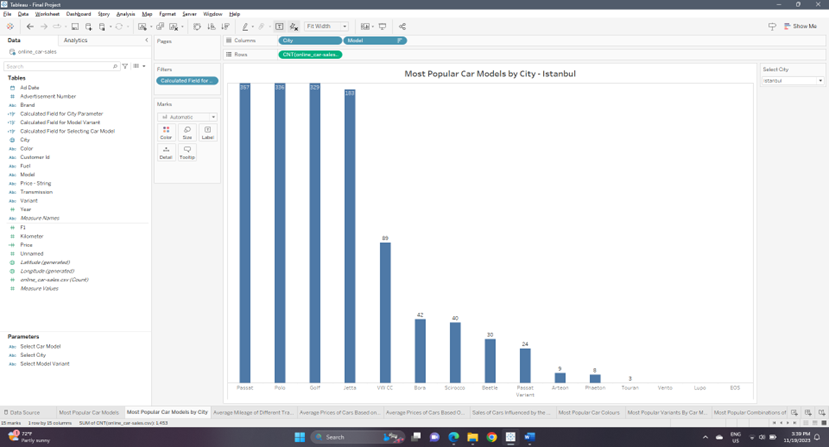
The plot used to address this question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



**2.** **Which cities have the most customers that purchase cars? What are the most popular car models there? Which cities have the fewest customers?**

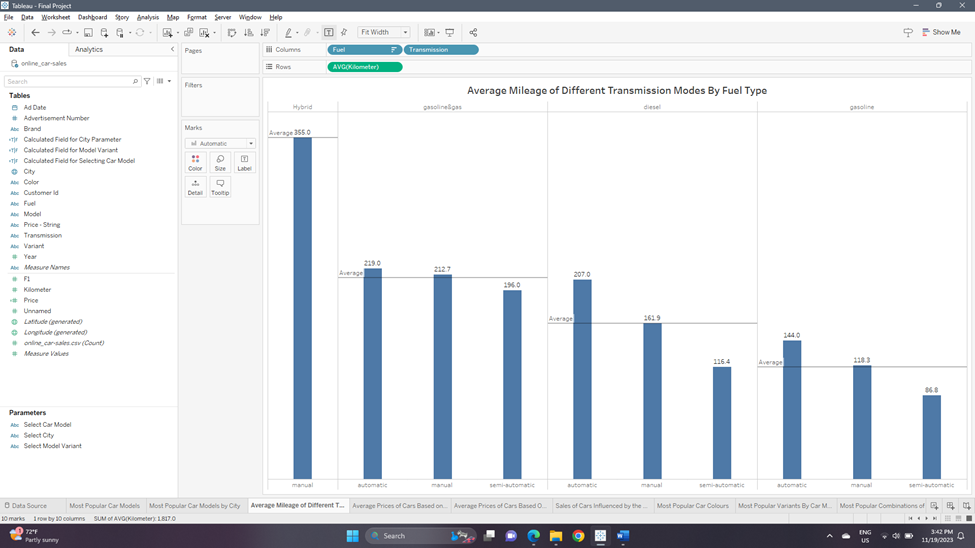
The plots used to address both the questions are **bar charts** as it is the best choice to display the number of quantities belonging to different categories. We are also implementing a user-input parameter to specify the amount that will colour the bars based on the input provided. The pre-attentive attribute used for the first chart is **length**, and for the second chart is **length and color** (to highlight the cities that had customers purchasing more cars than specified user input). Below are shown how the plots will look like:





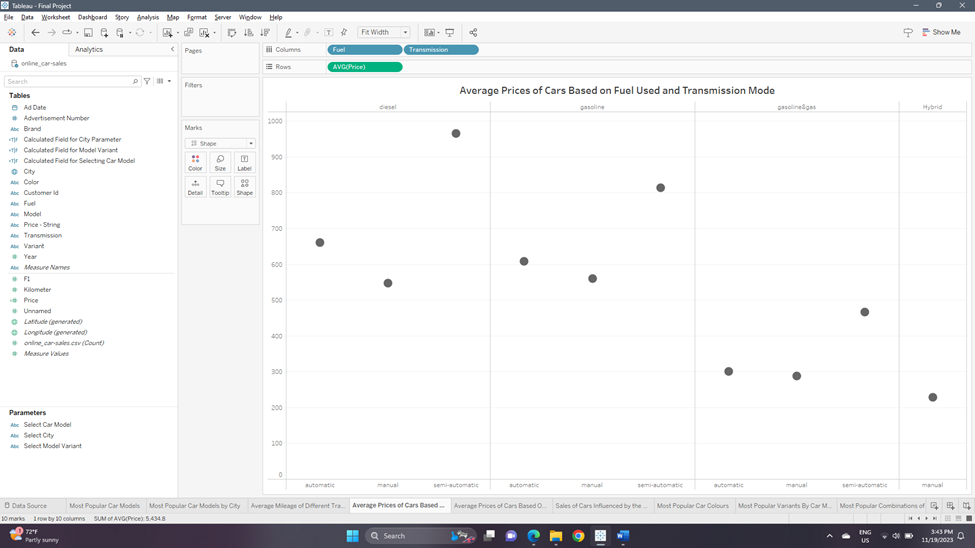
**3. What are the average distances traveled by cars that use different transmission modes and fuel types?**

The plot used to address this question is a **multi-line bar chart** as it is the best choice to display the amounts of quantities belonging to different categories of multiple attributes. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



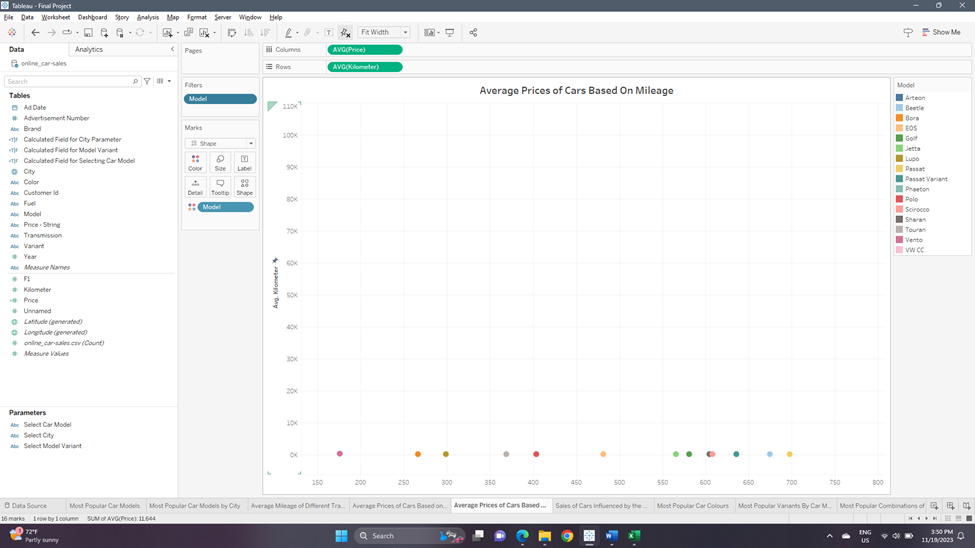
**4.** **How does the average price of gasoline cars compare to diesel and hybrid cars?**

The plot used to address this question is a **scatter plot** as it is the best choice to compare the amounts of quantities belonging to different categories or numerical attributes. The pre-attentive attribute used for this chart is **position**. Below is shown how the plot will look like:



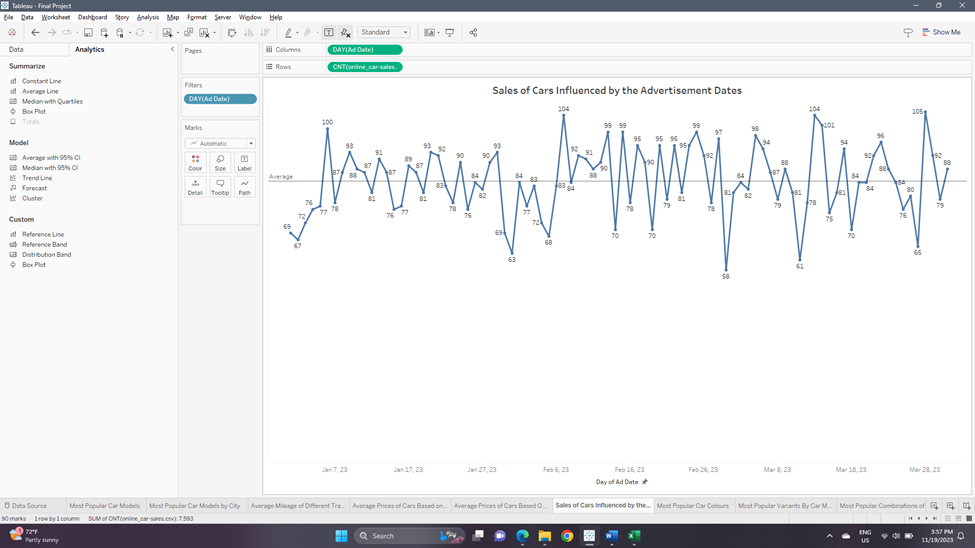
**5.**  **How does the average price of the different car models compare to their respective mileages?**

The plot used to address this question is a **scatter plot** as it is the best choice to compare the amounts of quantities belonging to different categories or numerical attributes. The pre-attentive attributes used for this chart are **position and color** (for representing different car models). Below is shown how the plot will look like:



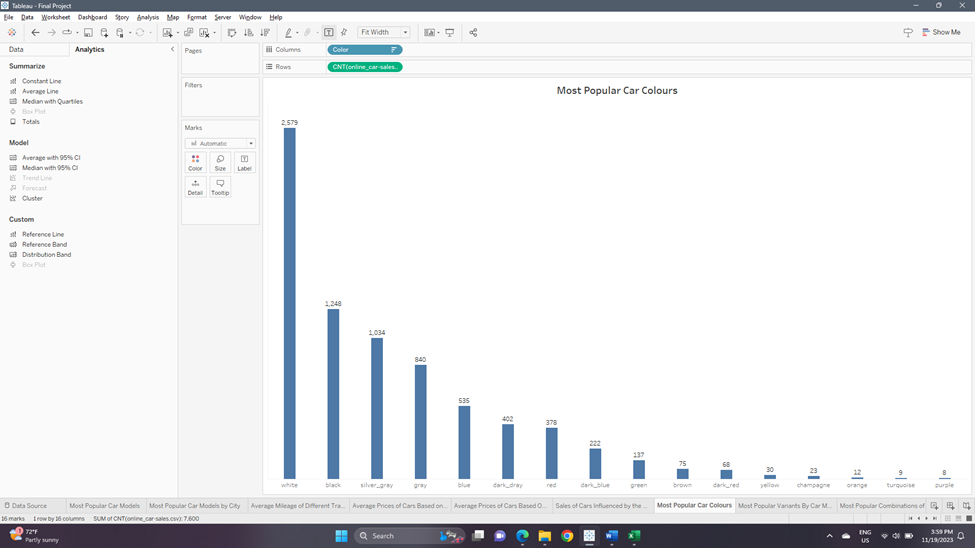
**6. Which month of advertisement dates had the highest effect on the sales of cars?**

The plot used to address this question is a **line chart** as it is the best choice to compare the amounts of quantities changing over a period of time. We have also added a reference line in the plot that denotes the average sales achieved over the whole time period. The pre-attentive attributes used for this chart are **position and orientation (of the line between consecutive points)**. Below is shown how the plot will look like:



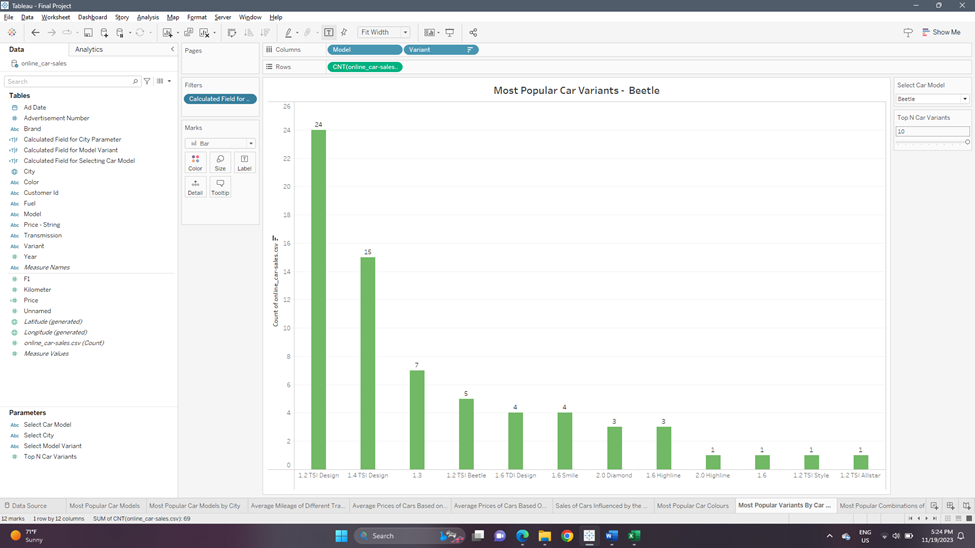
**7.Which car colors are the most popular among customers?**

The plot used to address this question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



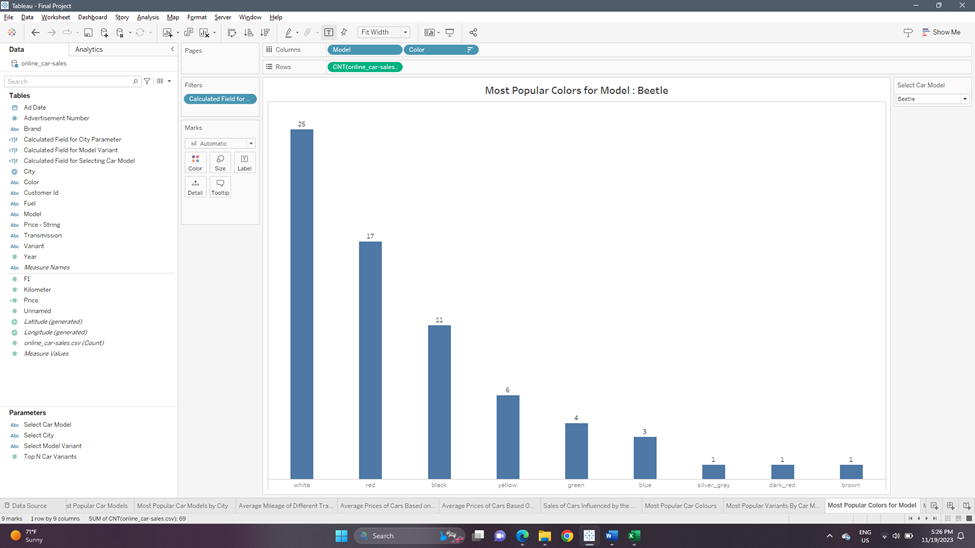
**8. What are the most popular variants of the different car models?**

The plot used to address this question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. We have added two parameters in this visualization; one for the user to select car model, and another is a ‘Top-N filter’ to give the user the option to specify how many top model variants should be displayed. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



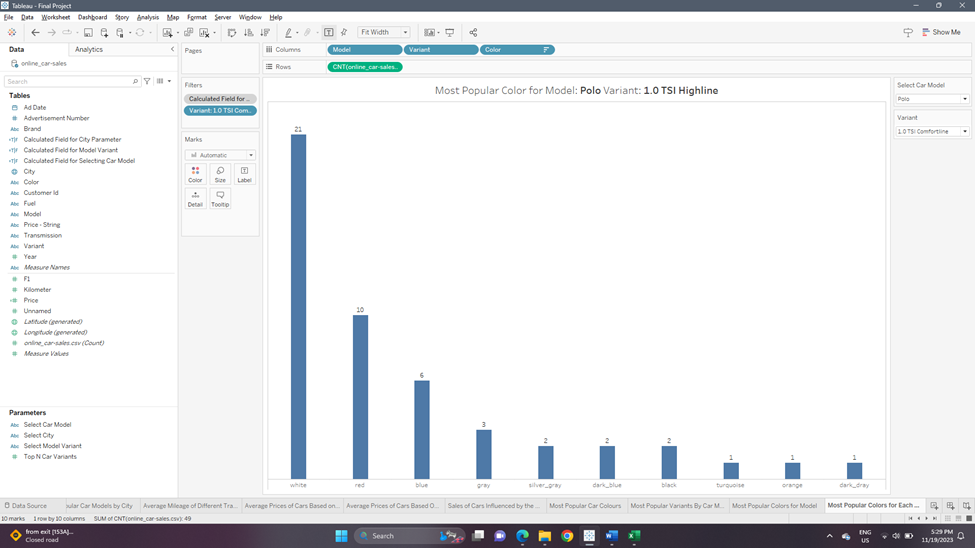
**9.** **What are the most popular colors for each car model?**

The plot used to address this question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. We have added one parameter in this visualization for the user to select car model for which the distribution of colours will be displayed. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



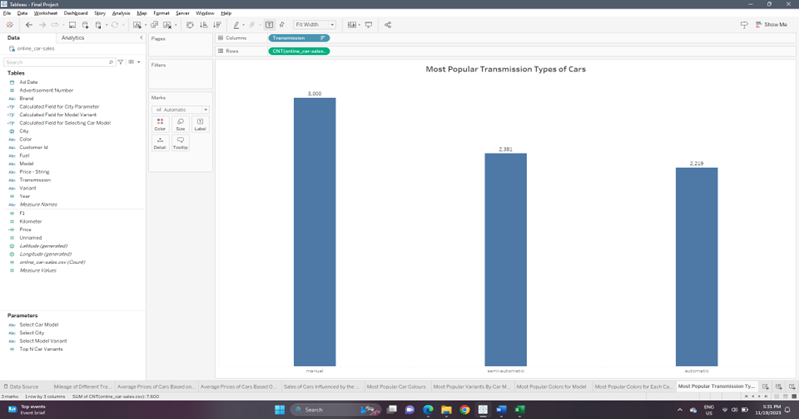
**10.** **What are the most popular colors for each variant of a car model?**

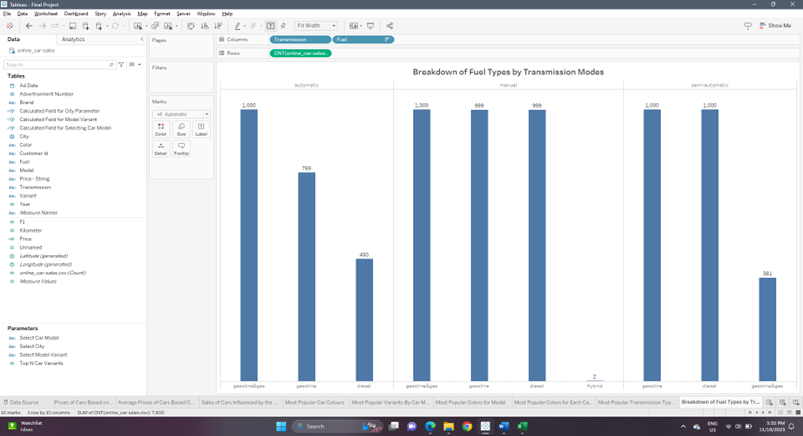
The plot used to address this question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. We have added two parameters in this visualization; one for the user to select the car model, and another to select the model variant. The pre-attentive attribute used for this chart is **length**. Below is shown how the plot will look like:



**11.** **What are the most popular transmission types in cars purchased? What are the counts of fuel types used by the car by different transmission types?**

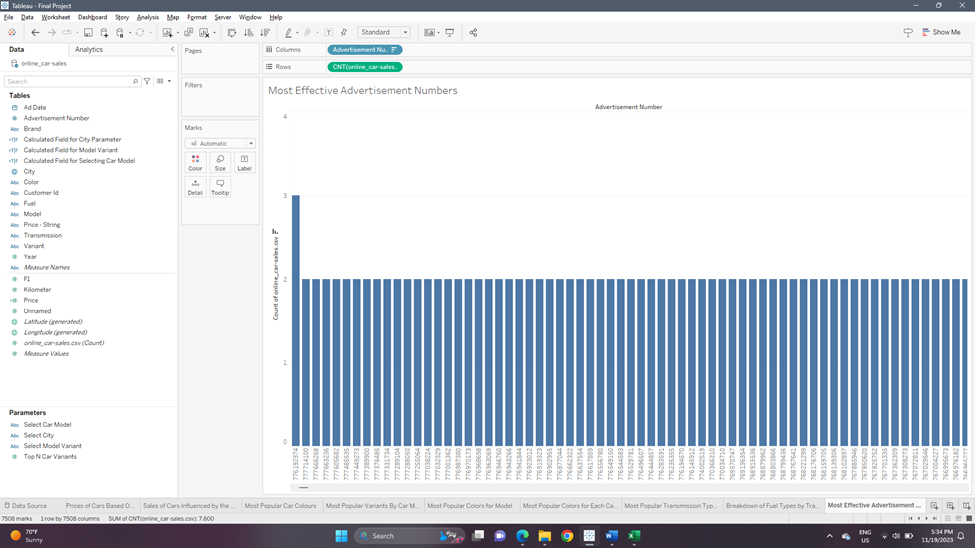
The plot used to address the first question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. The pre-attentive attribute used for this chart is **length**. The plot used to address the second question is a **multi-line bar chart** as it is the best choice to display the amounts of quantities belonging to different categories of multiple attributes. The pre-attentive attribute used for both charts is **length**. Below are shown how both the plots will look like:





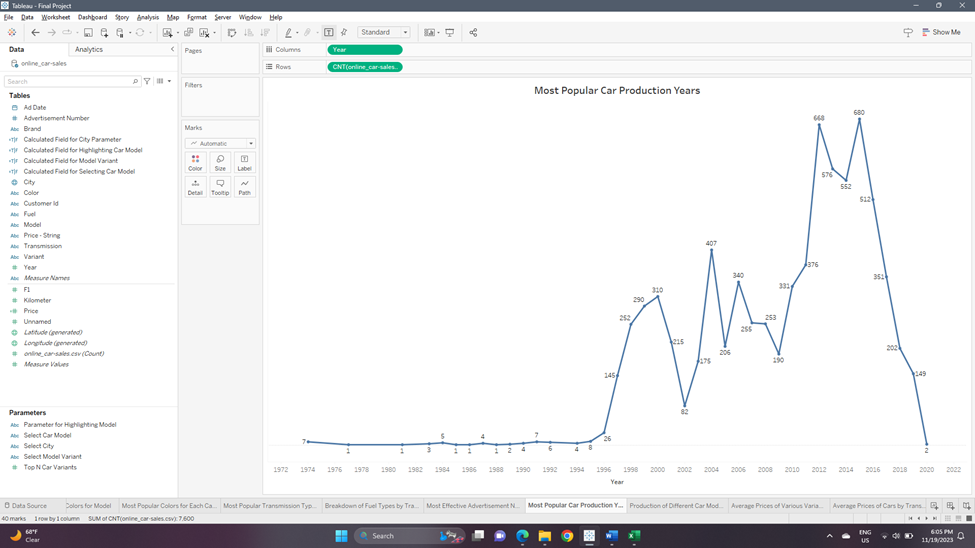
**12.** **Which advertisement numbers brought in the highest number of car sales?**

The plot used to address the question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. The pre-attentive attribute used for this chart is **length**. Below are shown how both the plots will look like:



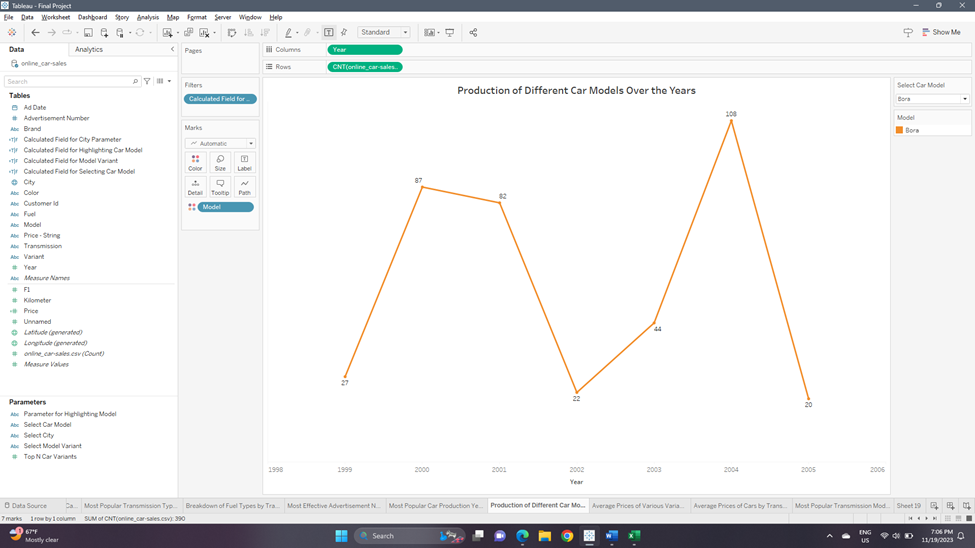
**13.** **Which year of production of cars is the most popular among customers?**

The plot used to address this question is a **line chart** as it is the best choice to compare the amounts of quantities changing over a period of time. The pre-attentive attributes used for this chart are **position and orientation (of the line between consecutive points)**. Below is shown how the plot will look like:



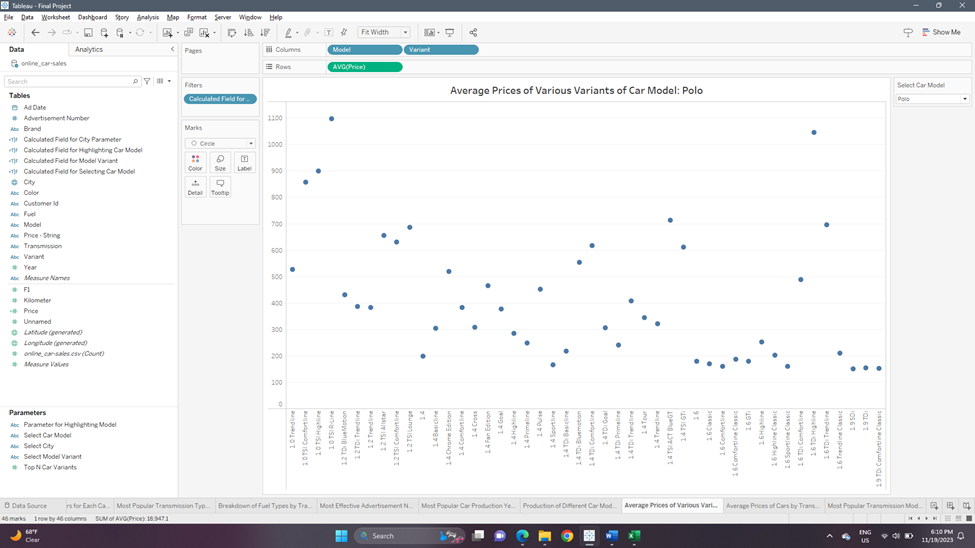
**14.**  **What are the distributions of years of production of different car models?**

The plot used to address this question is a **multi-line or a line chart** as it is the best choice to compare the amounts of quantities changing over a period of time. We have added one parameter in this visualization, for the user to select car model for which the amounts of different years of production will be displayed. The plot will be a multi-line chart when the option ‘ALL’ is selected that displays the values for all models. If a single model is selected, it will be a single line chart. The pre-attentive attributes used for this chart are **position, orientation (of the line between consecutive points) and color** (to represent different car models). Below is shown how the plot will look like:



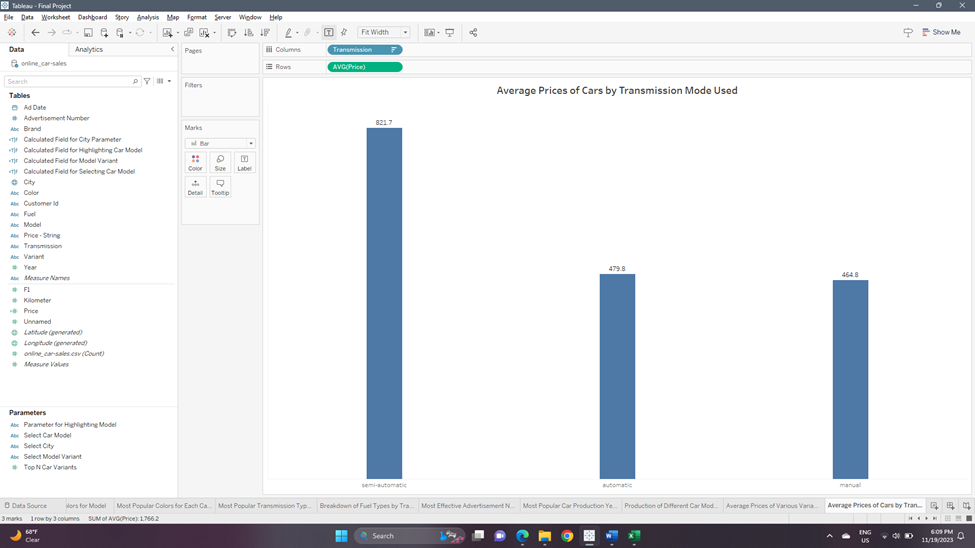
**15.**  **What are the average prices of the different variants of the same car models?**

The plot used to address this question is a **scatter plot** as it is the best choice to compare the amounts of quantities belonging to different categories or numerical attributes. We have added one parameter in this visualization, for the user to select the car model. The pre-attentive attribute used for this chart is **position**. Below is shown how the plot will look like:



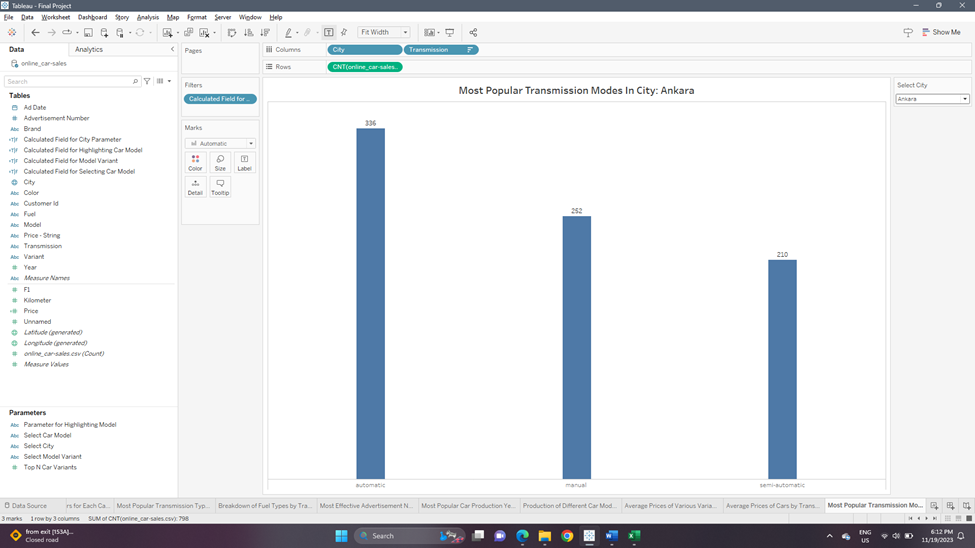
**16.**  **What are the averages prices of cars based on types of transmission mode used by the car?**

The plot used to address the first question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. The pre-attentive attribute used for this chart is **length**. Below are shown how both the plots will look like:



**17.**  **Which transmission modes are the most popular in each city?**

The plot used to address the first question is a **bar chart** as it is the best choice to display the amounts of quantities belonging to different categories. We have added one parameter in this visualization, for the user to select the city for which the distribution of car transmission modes will be displayed. The pre-attentive attribute used for this chart is **length**. Below are shown how both the plots will look like:



**Section 6: Interactivity**

We are implementing various interactive controls that will aid the user in better understanding the visualizations and are connected to various visualizations as required. Below is the list of interactive controls we have implemented -

* **Select Car Model –** This is a simple dropdown list that is used to specify the model of the car wherever required, and it is added in many visualizations. It is controlled using the calculated field ‘Calculated Field for Selecting Car Model’ that helps in filtering data based on the model selected. This parameter control is used in visualizations that answer the following questions – ‘What are the most popular variants of the different car models?’, ‘What are the most popular colors for each car model?’, ‘What are the most popular colors for each variant of a car model?’, ‘What are the distributions of years of production of different car models?’, and ‘What are the average prices of the different variants of the same car models?’. The values for this parameter are loaded from the ‘Model’ attribute of the dataset. It is a string attribute and hence has no range values.
* **Select Model Variant –** This is a simple dropdown list that is used to specify the variant of a particular car model wherever required, and it is added in one visualization. It is controlled using the calculated field ‘Calculated Field for Selecting Model Variant’ that helps in filtering data based on the variant selected. This parameter control is used in only one visualization that answers the following questions – ‘What are the most popular colors for each variant of a car model?’. The values for this parameter are loaded from the ‘Variant’ attribute of the dataset. It is a string attribute and hence has no range values. Another important property of this parameter is that it has been configured to display only relevant values based on the car model that has been selected using the ‘Context’ feature provided in Tableau. This helps in displaying only the variants that are particular to a specific car model.
* **Select City –** This is a simple dropdown list that is used to specify the city where the car was purchased wherever required, and it is added in two visualizations. It is controlled using the calculated field ‘Calculated Field for City Parameter’ that helps in filtering data based on the city selected. This parameter control is used in visualizations that answer the following questions – ‘Which cities have the most customers that purchase cars? What are the most popular car models there? Which cities have the fewest customers?’ and ‘Which transmission modes are the most popular in each city?’. The values for this parameter are loaded from the ‘City’ attribute of the dataset. It is a string attribute and hence has no range values.
* **Top-N Car Variants –** This is a slider provided for the user to specify the number of most popular variants of a particular car model that should be displayed as required. This feature is added to one visualization that answers the question – ‘What are the most popular variants of the different car models?’. This parameter only allows Integer values, with a range from a minimum of ‘1’ to a maximum of ‘10’ with a step-size of ‘1’.
* **Top-N Popular Models in Cities –** This is a slider provided for the user to specify the number of most popular car models that should be displayed as required. This feature is added to one visualization that answers the question – ‘Which cities have the most customers that purchase cars? What are the most popular car models there?’. This parameter only allows Integer values, with a range from a minimum of ‘1’ to a maximum of ‘10’ with a step-size of ‘1’.
* **User Input Amount for Cars Purchased –** This is a ‘Type-In’ input parameter created to allow the user to specify a particular number of cars purchased in a city. This parameter control is added to the visualization that answers the question – ‘Which cities have the most customers that purchase cars?’. The value provided by the user is compared with each value for a city that is used to create a bar, and then a reference line is added at that particular value. The bars of cities that have a higher number of cars purchased are then colored green, and the remaining bars are colored gold. A calculated field ‘Bar Color Based on Cars Purchased User Input’ is created and added as a filter to the visualization to make it work as intended. This parameter only allows Integer values, with a range from a minimum of ‘10’ to a maximum of ‘1500’ with a step-size of ‘10’.

**Section 7: References**

**Link to Dataset –**

<https://www.kaggle.com/datasets/bimervos/online-automotive-sales-statistics-volkswagen>

**Link to Mural –**

<https://app.mural.co/t/dvassignment0284/m/dvassignment0284/1699563981327/42f506657e2cd9915e45641bf72383a216df88c2?sender=ua05eef829bc2a8f507a20491>

**Link to Published Dashboards –**

**Dashboard 1 –** <https://public.tableau.com/app/profile/prithvi.bhatt/viz/DVFinalProject-Group8/Dashboard1?publish=yes>

**Dashboard 2 –** https://public.tableau.com/app/profile/prithvi.bhatt/viz/DVFinalProject-Group8/Dashboard2?publish=yes

**Dashboard 3 –** https://public.tableau.com/app/profile/prithvi.bhatt/viz/DVFinalProject-Group8/Dashboard2?publish=yes

**Dashboard 4 –** https://public.tableau.com/app/profile/prithvi.bhatt/viz/DVFinalProject-Group8/Dashboard4?publish=yes