Attila Serfőző- Semester Project Documentation

1 Overview

I decided to create a visualization for the Ford Focus used cars available on Hasznaltauto.hu. The motivation behind is that I also have a 2012 Ford Focus Hatchback car, thus I was interested in the distribution of these vehicles across different attributes to benchmark the value and condition of my car based on the available data on the website.

2 Dataset

The dataset I used was scraped from hasznaltauto.hu, the largest Hungarian used cars website, on 11th of March 2021 by an R-code I written myself. I collected all the data on 947 used cars and wrote another R-code for data cleaning and variable selection. I selected ID, name, price in HUF, registration year, age (calculated from 2021 January) in years, condition, type, kilometers, color, fuel type, transmission type, cylinder capacity, performance in horsepower and kilowatts variables for visualization purposes.

During the transformation-preparation exercise I created numerical variables as many included the metric in their value, thus was identified as string. I got rid of cars with missing price values and cars with rare categories having only one or two observations like cabrio cars, electric/gas cars, or sequential transmission vehicles. Finally, I unified observations in category variables like type, transmission, color, name and equipment to reduce variation due to different wording on the website.

3 Dashboards

3.1 Purpose

The goal of the first dashboard is to show distribution of available Ford Focuses across different categories and attributes. The main messages of the first dashboard include:

- Most of the Ford Focuses on hasznaltauto.hu are priced between 2-4 million HUF and are uploaded to the website after passing 100 000 km.
- Cars with an Excellent or Novel condition running less than 20k kilometers per year can be a good acquisition.
- Using the car type filters on the top the user can see that while 2/3 of Hatchback and Sedan cars run with Petrol, on the other hand 2/3 of Estate cars are diesels.
- The tree map highlights for us that nowadays 3/4 of the cars have conservative colors (grey, sliver, white, black).

The goal of the second dashboard is to show the connection of car prices with the different attributes. The main messages of the second dashboard include:

- Cars after leaving the salon are quickly losing from their value in the first 2-3 years and/or 100 000 kilometers, the trend line highlights the price movement.
- Most expensive engine and equipment types are dominated by the sport versions and Titanium package equipped cars which contains most of the extras.
- On the boxplots the user can suspect that petrol and manual transmission cars have a larger variance in their prices with some extreme pricing. The reason behind is that sport

version of Ford Focuses (RS, ST2, ST3 cars) can only be ordered with petrol and manual transmission and their pricing is always high as they are rare.
- Automatic transmission and extra colors usually cost an extra.

3.2 Visual components

First dashboard:
- The main element is the bar chart on the top left corner. I decided to use bar chart, as it can highlight very effectively the skewness of some variables like the price variable or avg. kilometers per year.
- For showing the car conditions I decided to use a table, as on hasznaltauto.hu there are no clues about the ranking of the condition attributes. I found it a huge problem, as people sometimes randomly categorize their vehicle based on this field. In the table I wanted to provide details of these categories so the user can get some impression on which categories are the best to look for.
- I selected the 100% horizontal bar chart to visualize the distribution of fuel type and transmissions, as it makes comparison easier when we use the car type filter on the top.
- I decided to select a tree map for the color distribution, as I wanted to highlight the vast majority of conservative colors nowadays on the road. I decided to use the exact colors for coloring, as I think it show the message better than just using the different shades of blue. To avoid misleading the attention I avoided to use too intensive tones of these colors.

Second dashboard:
- The main element is the line chart on the top left corner. I decided to use it, as it highlights the price decrease/increase trend effectively.
- I selected the boxplots to visualize the distribution of car prices by fuel type and transmission type, as it provides more detailed information than a bar plot with the average prices of these categories. The user can get more detail in which cases they can expect wider scale of prices (petrol or automatic) or extreme valuations (petrol or manual) or higher prices (petrol or automatic).
- I decided to select the horizontal bar plot for the average price per color chart, as it can easily show using the colorings that more colorful cars can be more expensive maybe as they are rare nowadays.

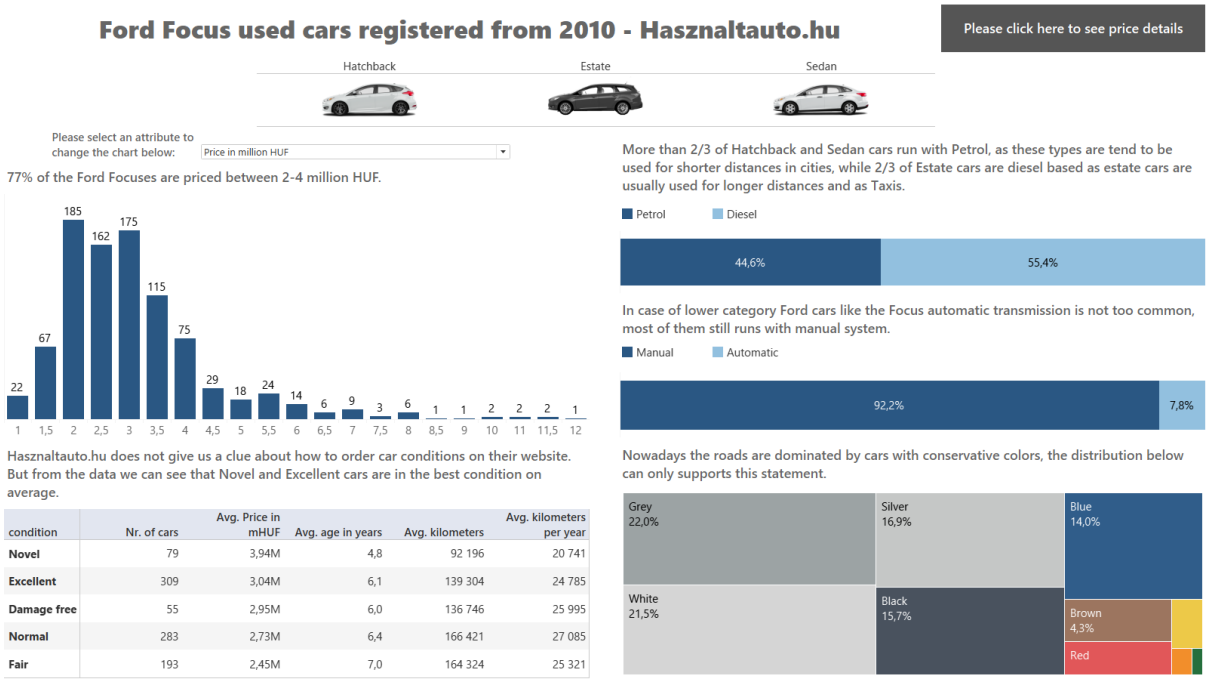
3.3 Functionality / Tableau techniques / interactivity

On both dashboards I created the main car type filter using ford focus PNG files. The user can navigate between the two dashboards using the button in the top right corner.

On the first dashboard the main challenge was the bar plot, as here I needed to combine different tableau techniques. I used a parameter selector to make attribute selection available, but first I needed to create bins with tableau calculations from the desired variables. Unfortunately, it turned out that results of the basic tableau bin or group creator, cannot be used in tableau calculations therefore in the parameter selector so I found it the hardest to solve this issue. I also created flexible title for this chart which means as the user changes the parameter, the descriptive title changes according to it as well. The first dashboard can be filtered with all the charts except the table as there I did not find it necessary.

On the second dashboard the line chart is using a parameter selector and the title is changing according to the parameter similarly to the first dashboard. I also added a trend line from the analytics pane to highlight the price movement trends. The Top 10 most expensive car type chart includes context filters to avoid missing out values due to filtering first the top 10 and applying then the other filters. The user can also decide to include/exclude sport cars from the top 10, as they are always the most expensive. The second dashboard can be filtered with the fuel and transmission type charts, colors and the car type pictures on the top.

4 Dashboard screenshots



condition	Nr. of cars	Avg. Price in mHUF	Avg. age in years	Avg. kilometers	Avg. kilometers per year
Novel	79	3,94M	4,8	92 196	20 741
Excellent	309	3,04M	6,1	139 304	24 785
Damage free	55	2,95M	6,0	136 746	25 995
Normal	283	2,73M	6,4	166 421	27 085
Fair	193	2,45M	7,0	164 324	25 321