VIETNAM NATIONAL UNIVERSITY – HCM

**INTERNATIONAL UNIVERSITY**

A blue and white logo

Description automatically generated

**Project Proposal**

**Topic: The development of interactive cars data visualization**

**Course: Data Science and Data Visualization   
 – IT138IU**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Name** | **Student ID** | **Contribution** |
| **1** | **Huỳnh Minh Duy** | **ITCSIU21174** |  |
| **2** | **Nguyễn Nguyên Bảo Phú** | **ITCSIU21216** |  |
| **3** | **Tạ Trung Hiếu** | **ITCSIU21180** |  |
| **4** | **Trương Trí Dũng** | **ITCSIU21126** |  |
| **5** | **Nguyễn Đức Minh** | **ITDSIU22148** |  |

**Assoc.Prof Tran Thanh Tung**

**Department: School of Computer Science and Engineering**

1. **Introduction**
   1. **Project Motivation**

During our second-year final semester, we - five people had the opportunity to participate in the university's open tour Bosch workshop. As we wandered through the halls of a renowned car manufacturing plant, the symphony of machines echoed our curiosity. We witnessed the meticulous assembly of vehicles, each part seamlessly integrating to form the embodiment of modern engineering.

The passion for cars grew as we spoke with experienced engineers who recounted stories of tenacity and triumphs in the automotive sector. From the advancement of hybrid technology to the complexities of aerodynamics, we were attracted by the infinite possibilities of automobile design. Since then, we have developed a curiosity for the different types of interactive cars and do some bit of research into them.

* 1. **Project Objectives**

The main queries that the visualization may be able to address are what kinds of automobiles a certain brand typically produces and the relationship between the cars' prices and other characteristics.   
Users of the visualization may compare items from different automakers and evaluate differences in terms of pricing, branding, and specifications. Enthusiasts of automobiles might investigate the specifics of individual components and their relationships.  
Those who want to purchase automobiles can have a better grasp of the specializations and price ranges of various car brands.

- Some questions that can be answered by using the visualization:

+, Are diesel cars generally more expensive than gas cars?

+, Is there any correlation between engine size and horsepower?

+, Does the length of cars affect the curb-weight?

* 1. **Project Schedule**

Week 1-3:

* Choose a visualization project based on research interests and motivations
* Choose a relevant dataset, sketch visualization ideas, and analyze the dataset.
* Write Proposal

Week 4-6:

* Review web programming, learn D3.js library and visualization theory

Week 7-8:

* Start coding to build the overall structure of the visualization.
* Write Progress Report

Week 9:

* Add detail improvement to the visualization.

Week 10:

* User-ability testing and evaluation, finish report.

Week 11-12:

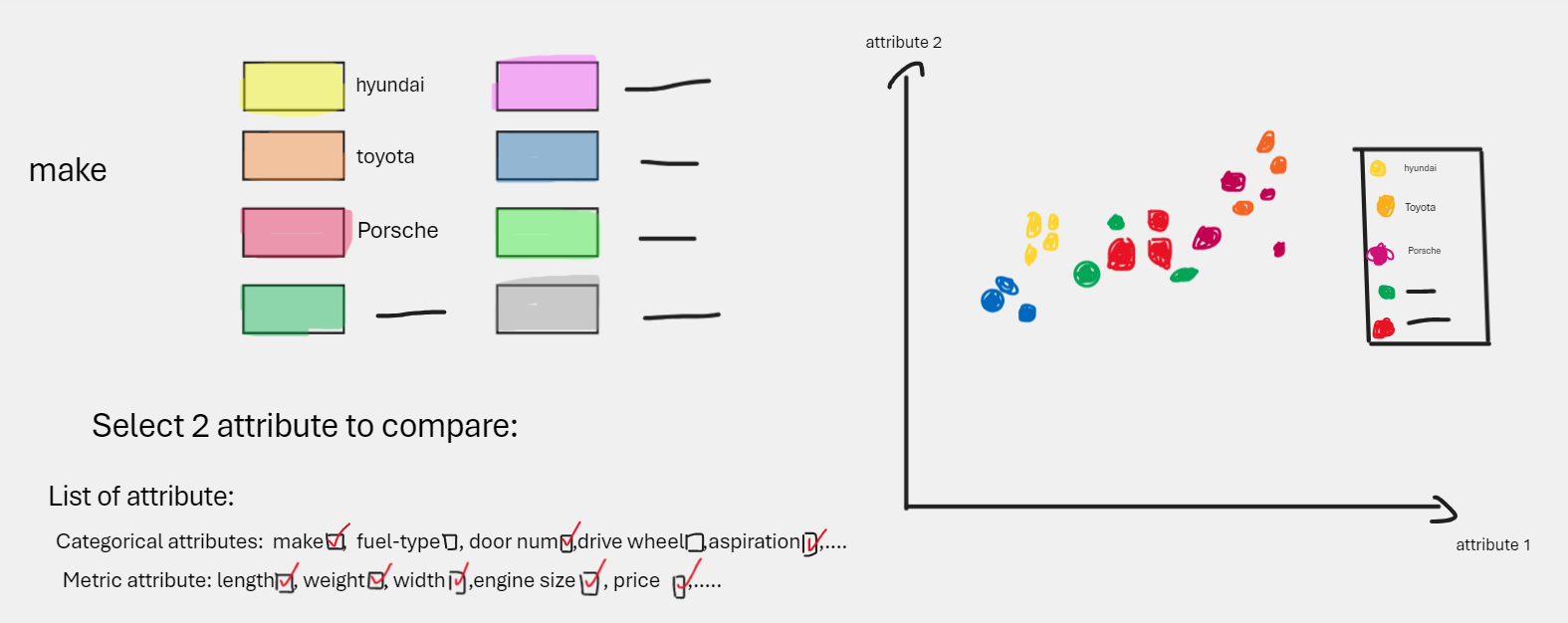
* Prepare for presentation.

1. **Data Source**

The dataset we proposed for this project is called Automobile Dataset which is accessible for use at the UCI Machine Learning Repository (<https://archive.ics.uci.edu/dataset/10/automobile>). As you download the dataset from the link, you can observe a comprising table stored in CSV file, the dataset encompasses 205 distinct car models. Its attribute features span categorical, integer, and real types. Notably, there exist 10 categorical variables (such as fuel type and number of cylinders) and 14 continuous variables (including price and length), alongside indicators for risk assessment (symbolling) and normalized losses.

1. **Visual Data Sketches and Ideas**

This visualization we vision is designed to help people compare various car companies. With the layout illustrated in Figure 1, users can easily compare different makes and see how two attributes correlate. You can choose the makes and attributes you're interested in, and a scatter plot will appear. Furthermore, you have the flexibility to select or deselect makes and change attributes smoothly. Hovering over data points on the scatter plot reveals more detailed information at the mouse's location.

****

*Figure 1: Visual Idea 1*

This design lets you check out how many cars fall into specific categories. For example, we divide prices into various ranges, and then you can see how many cars fit into each price range through a bar chart, as illustrated in Figure 2. You can hover over each bar to see a pie chart that breaks down the percentage of cars based on their makes.

**A graph with a pie chart and a pie chart

Description automatically generated**

*Figure 2: Pie chart and Bar chart*

In the Figure 3 below presents a parallel coordinate plot, where every line stands for a data point, with colors denoting the different car makes. Users have the ability to pick a range along an axis, which then highlights all cars possessing attribute values within that chosen range. This layout empowers users to simultaneously view numerous attributes and investigate correlations between variables by selecting value ranges. Additionally, the colors aid in distinguishing between car makes.

**A diagram of a graph

Description automatically generated**

*Figure 3: Parallel coordinate plot*

1. **Questions to clarify the requirements**

**Q1:** The dataset and data visualization we found was done by others, however, their visualization was not optimized and their algorithm was not a good way to approach visualizing data. So our question is instead of making a new data visualization, can we focus on improving the way to approach that visualization and make more visualization charts, detail tables and plots based on their existing data?