UNIVERSITY OF DELAWARE
LERNER BUSINESS & ECONOMICS

Data Analysis Project (Phase 1)

Team members' full names	Nhinhi Nguyen and Steven Pham

Introduction , project purpose, the questions you're interested to answer, and etc. At least 6 lines with Font 12, Times New Roman.

The Excel starter file is loaded with information regarding motor vehicle specifics, such as the first sheet describing the cars' year, mileage, size, and more. The second sheet has details about the cars' VIN ID, fuel tank type, and volume. As we viewed the abundant amount of information we interpreted it as files from a car dealership organizing all of the cars it has in their stock. We want to organize the data to sort the vehicles into different categories. This will be helpful for potential buyers who could be families looking for a minivan, a newlywed couple looking for a sedan, and etc.

Questions:

- Which cars can fit a family of 5?
- Which car model has the best fuel economy? (categorize cars on tank volume)
- What vehicle model has the best-seller ratings?
- Can we categorize the cars into mileage categories? (Low, medium, high mileage)
- What cars have had previous accidents?

Data cleaning

Please explain what you did for data cleaning. If you used any formula, please copy and paste the formula after your explanation.

- **Vlookup formula1:**

```
=VLOOKUP(A2,Sheet2!$A$1:$C$100000,2,FALSE)
```

- **Vlookup formula2:**

```
=VLOOKUP(A2,Sheet2!$A$1:$C$100000,3,FALSE)
```

- **Split data in one column into two columns (explain which columns you split and how):**

We splitted BodyType_Exterior Color into two columns by using the text to columns function. We named one column "Car Type" and the other "Exterior Color".

- **Extract Year, Month, and Day from Date/Time (explain from which column/columns you extracted and how):**

We extracted the year from the listed_date by using =TEXT(X2,"yyyy") and renamed the new column "Listed Year".

- **Extract numeric part of the columns that have combination of number and text (Explain from which column/columns you extracted and how):**

We extracted the numeric value of the fuel tank volume to separate the number from the gallon, so that when we categorize cars for how much gas they can hold we have the numeric value to do so.

- **Remove duplicate records (explain how you removed duplicates):**

To remove duplicates we selected the entire dataset of sheet 1 and clicked on the data tab, then clicked on remove duplicates, selected all fields then pressed OK.

- **Remove rows with blank cells (explain how you removed rows with blank cells):**

We removed rows with blank cells by going to find & select, "go to special" and pressed "control -" and deleted entire rows.

- **If you did any other data cleaning, please explain here:**

Formulas for creating a new field (at least 3 formulas are required)

Please copy and paste your formulas here and explain them.

- **Nested If statements (provide the formula and explain what it does):**

=IF(AC2="2 seats","Single/Couple",IF(AC2="4 seats","Small family",IF(AC2="5 seats","Big family","")))

We created a nested if statement to see the maximum amount of seats a vehicle has and sorted them into categories. For instance, a car with a maximum seating of 2 is for a single/couple while a car that has 4 seats is designated for a small family. Then, a car with a maximum seat of 5 is for a bigger family which we stated that they were a "big family".

- **Combination of IF and AND (provide the formula and explain what it does):**

=IF(AD2<=30000, "Low Mileage",IF(AND(AD2 <= 80000, AD2 >=30000),"Medium Mileage",IF(AD2 > 80000,"High Mileage")))

We created a combination of an IF and AND statement to categorize the cars based on if they had low, medium, or high gas mileage. This is a useful category for potential customers who want to gauge what car to purchase based on its previous use.

- **Combination of IF and OR (provide the formula and explain what it does):**

=IF(OR(O2=TRUE,AD2>=150000),"Unreliable", "Reliable")

We created a combination IF and OR statement to highlight what cars have either been in an accident or have mileage over 150,000 miles. This will be helpful for customers who want to know these things about the car they will purchase.

=IF(E2<= 25, "Low City Fuel", IF(OR(E2 >= 30, E2 <= 35), "Medium City Fuel", IF(E2 > 36, "High City Fuel", "")))

```
=IF(Q64<= 25, "Low Highway Fuel", IF(OR(Q64 >= 30, Q64 <= 35), "Medium Highway Fuel",
IF(Q64 > 36, "High Highway Fuel", "")))
```