**Case Study - Hypothesis Testing for Data Science**

**Introduction**

Insurance broker Ankur has a dataset of car insurance claims. He has a few questions in his head, such as whether length and width have any relationship with one another that contributes equally to claiming the car insurance, in order to get insights from the dataset and understand what parameters of the dataset are impacting more on claiming insurance. Please assist him in understanding the dataset.

**Problem Statement**

Ankur and his friend want to perform hypothesis testing on the dataset to decide whether one hypothesis is true or false on the basis of different attributes. This analysis aims to check which attribute contributes to a particular hypothesis.

**Initial Analysis**

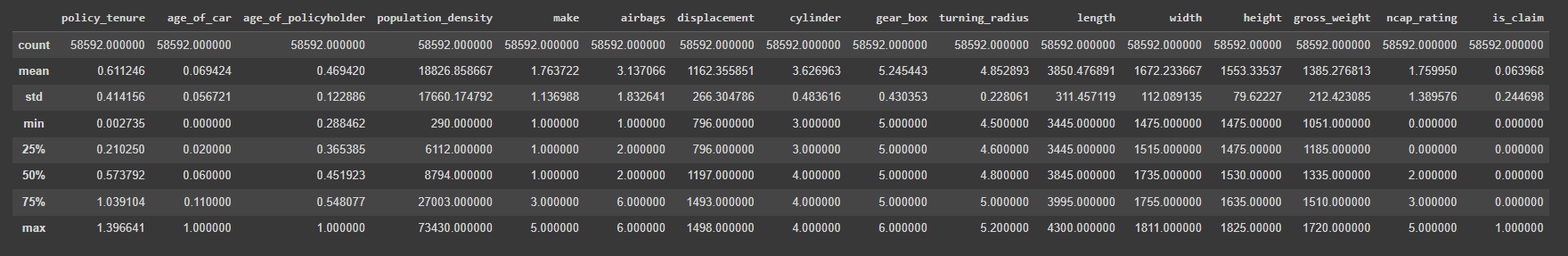
Get the dataset in the form of a table in Python and find the number of records and features present in the dataset.

**Example Dataset**

They have taken the following dataset for this case study: [Car Insurance Claim.csv](https://drive.google.com/file/d/1ztFeilEMGgEIEkAfPAplfDUIwqtLp_2z/view?usp=share_link)

You can see some of the records of the dataset shared with you and you can also learn about some of their features.

Table: Car Insurance Claim.csv



The df table captures all the names of the different car insurance features and the whole dataset.

**Case Study Questions**

1. Each of the following case study questions can be answered using Python:
2. What is hypothesis testing?
3. Load the CSV file into the table format and find the shape of the data frame.
4. Visualize the correlation between features.
5. Remove the unnecessary columns.
6. Check for the Null and Duplicate values present in the dataset and remove them.
7. Check if the average weight of a car in any group in this dataset is close to 1380.
8. Check if there is any difference in claiming insurance based on the different types of fuel.
9. Check whether the insurance claim depends on the different types of steering in vehicles?
10. Check whether the group means of different columns ('make' and 'ncap\_rating') are equal or not?
11. Check if checking whether the max torque and max power of a vehicle correlated or not.
12. Check if the insurance claim depends on the different types of segments of vehicles?
13. Check whether the length and width are associated with each other or not for claiming the car insurance?

**Next Steps**

Treat the outliers present in the dataset and transform the features in the dataset to get normally distributed features.