**EV Market Segment Analysis Report**

**Name: Yaswanth Kumar Puli**

**Problem Statement**

**Focusing on Consumer’s demographic profiles, knowledge, practice and attitude. How these factors are affecting Electric Vehicle Adoption in India. By conducting a survey, we got the data.**

**The data is classified into 2 segments**

**section - A focusing on consumer’s profile (Gender, Age, Occupation).**

**(Gender 1-Male, 2-Female), (Age [1] - 18-24, [2]- 25-31,[3] - 32-38**

**[4] - 38 years above), (Occupation [1]- Professor, [2]-CA, [3]-Research Scholar, [4]-Doctors)**

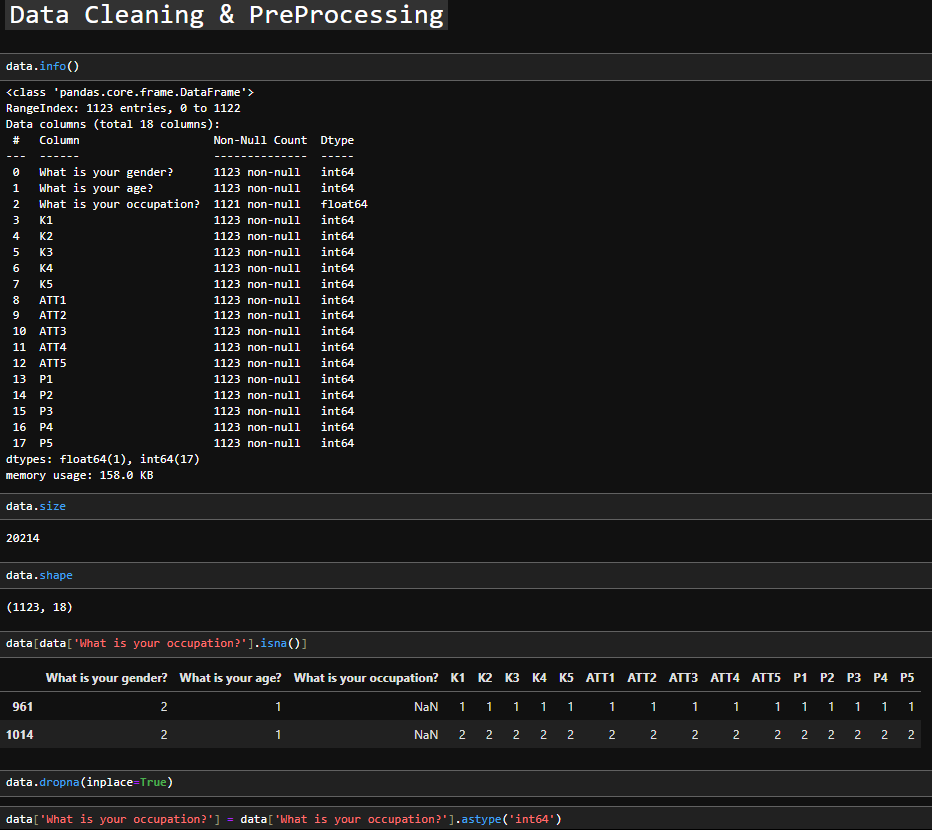
**Section – B focusing on knowledge(K-Segment), Attitude (ATT-Segment) and Practice (P-segment).**

**All these segments contain scale of 5Q each (k1-k5, ATT1-ATT5, P1-P5).**

**The Answers are scaled from (1 to 5, 1 being strongly agree, 2 – agree, 3 – neutral, 4 – disagree, 5- strongly disagree) for each question.**

**Dataset Overview**

**Performing Data Cleaning and Preprocessing**

****

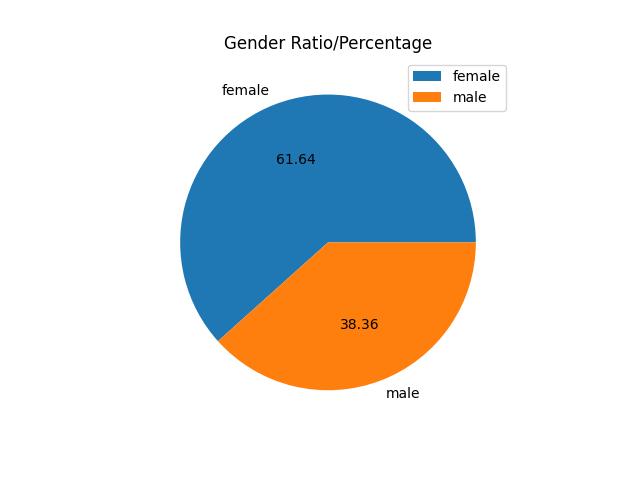
The dataset was almost clean, it had 2 null values and I dropped those rows. And did a type conversion from float to int to the occupation feature/column.

**Exploratory Data Analysis**

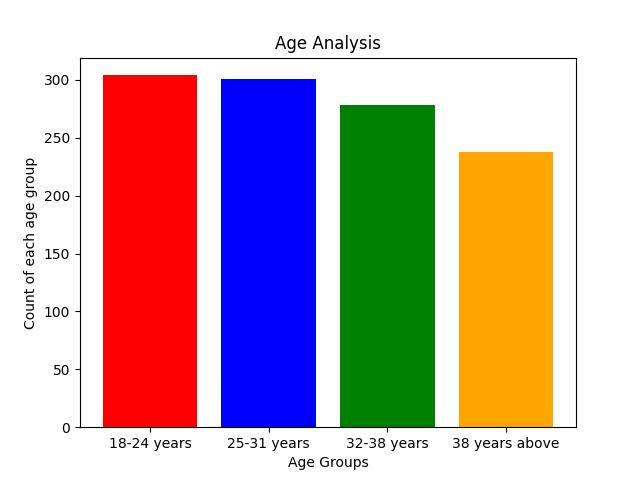
EDA typically refers to analyzing and visualizing through exploring the given data and get key insights, patterns and trends through the statistical analysis. Often the first step in the data science workflow before building models or making predictions.

**Section – A**

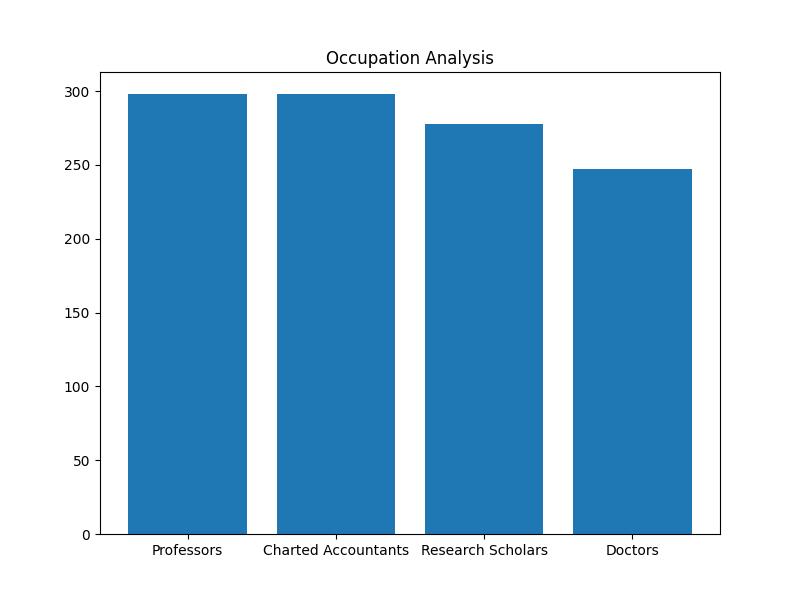
**Gender Analysis**

****

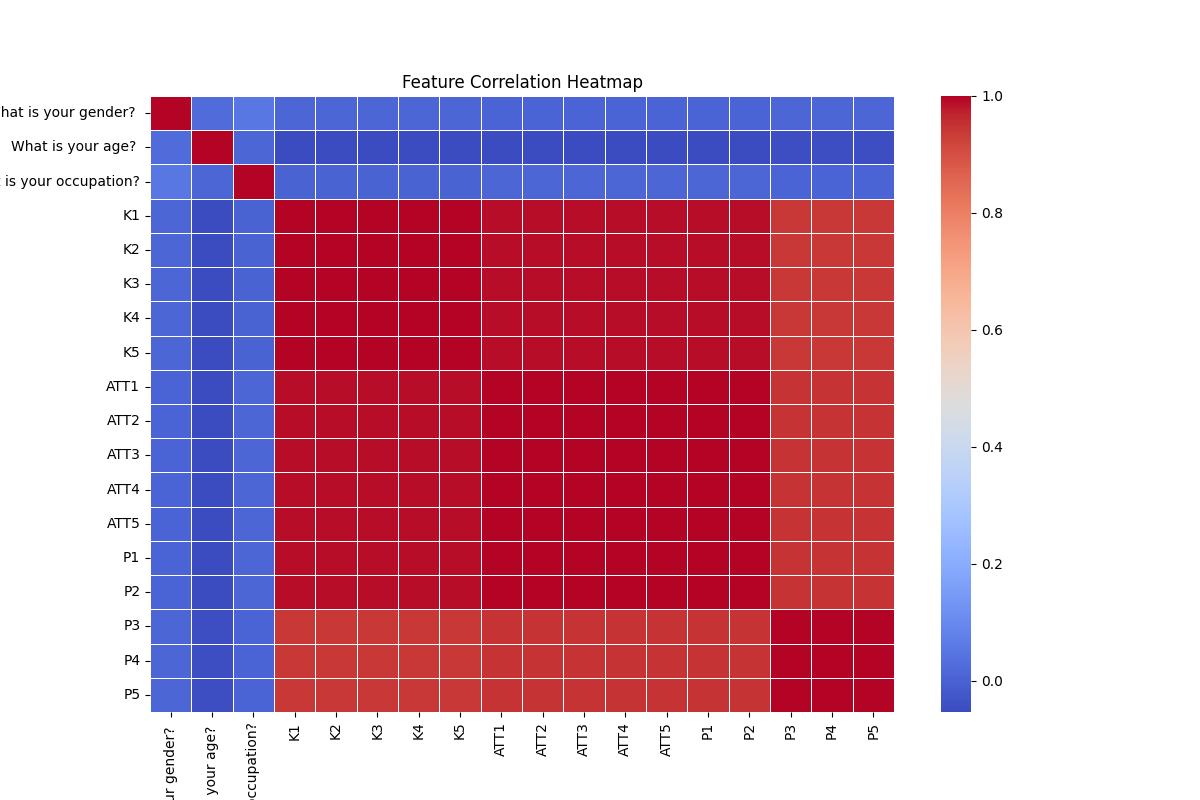
**Age Analysis**

****

**Occupation Analysis**

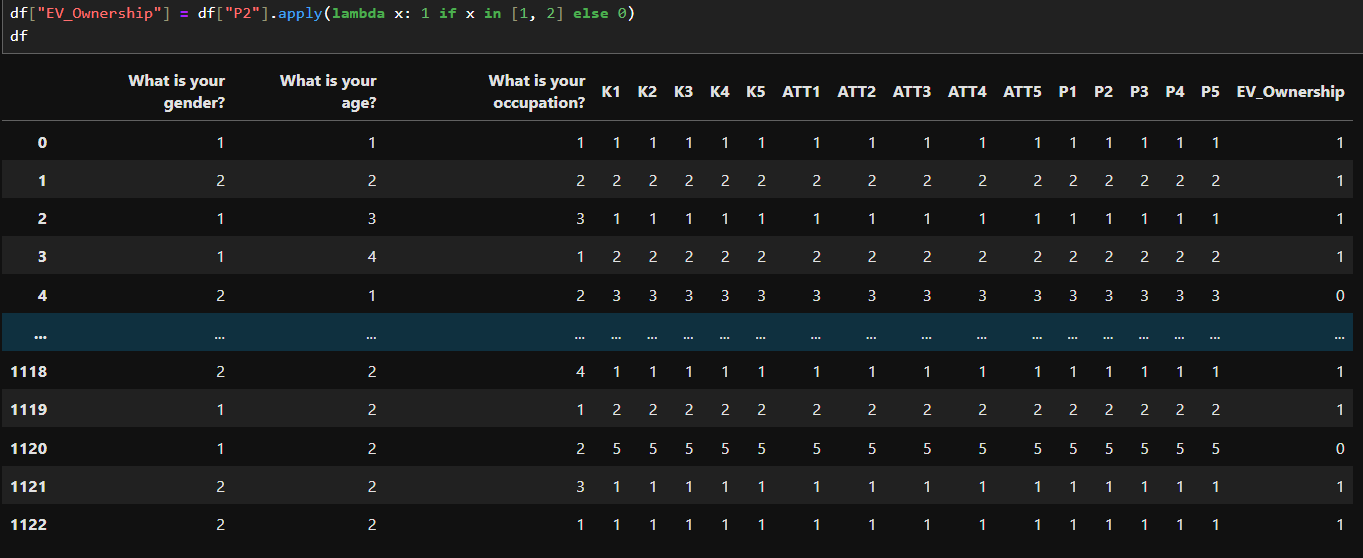
****

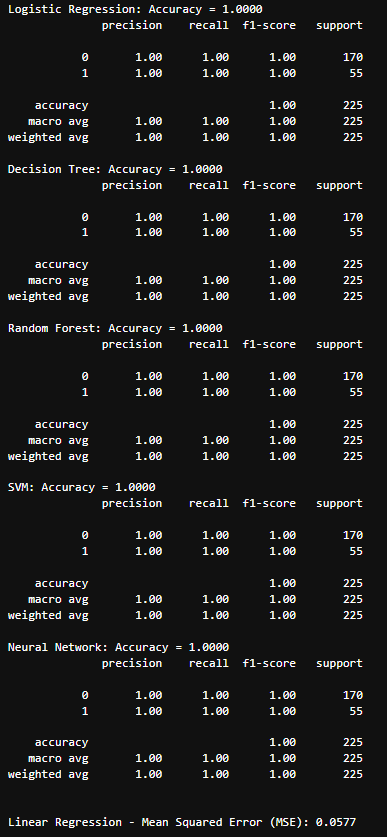
**Feature Correlation Map**

****

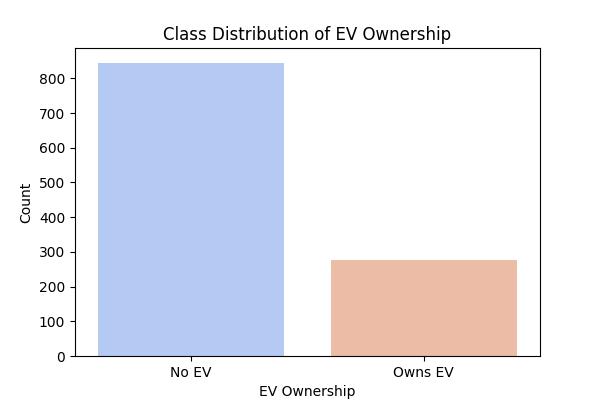
**ML Model’s Used in the project**

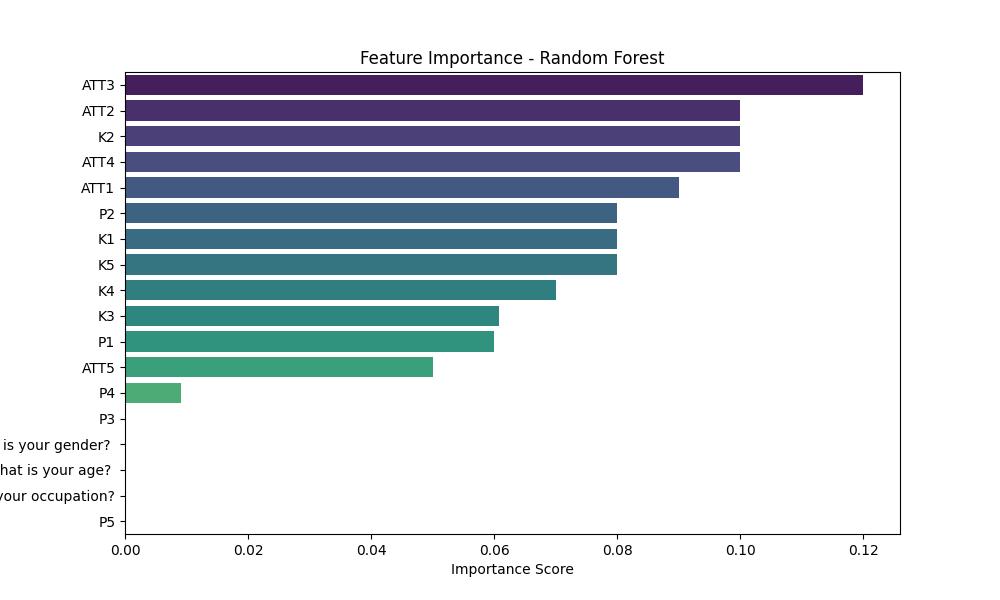
Firstly, I used classical machine learning models. Created a target variable ‘Ev\_Ownership’ using p2 variable (currently own or lease an electric vehicle), and used section-b features as Training parameters. Used multiple supervised learning predictive models to predict the ev\_ownership. (linear regression, logistic regression, decision tree, random forest, SVC) and created a classification report. And also, the visualization reports.

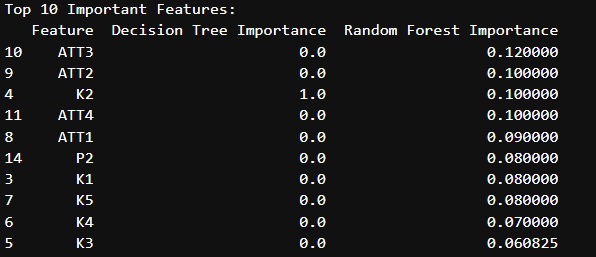


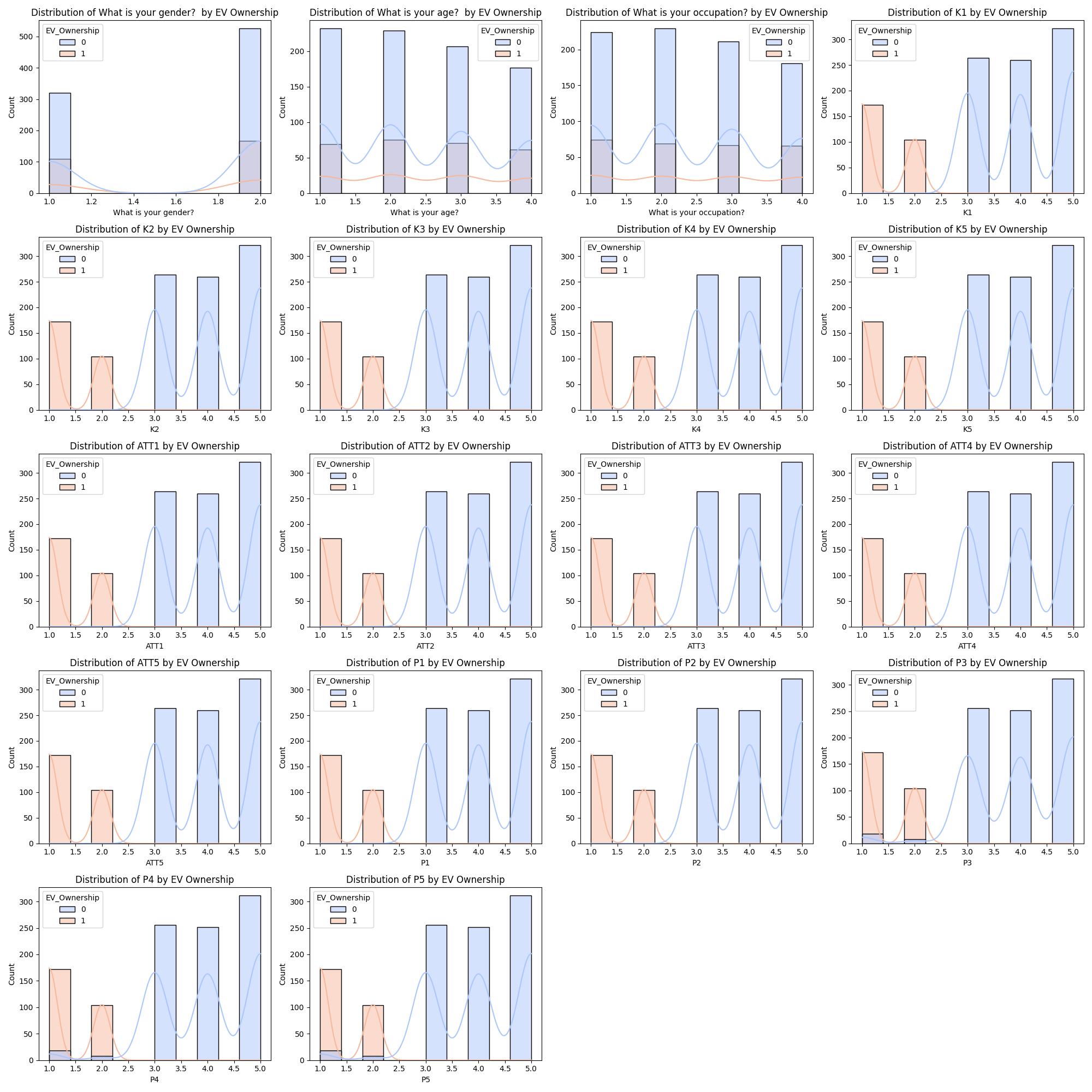


Used random forest for feature importance predicting top features that are useful. And ev ownership by class distribution.









**Final conclusion & insights gained from the research/analysis work**

**Insights:**

From section-A profiling, in gender analysis women population/ratio is greater than men, most of the people fall under the age group of 18-24 and 24-31 years, most people are in professors’ occupation followed up by CA and research scholars and least are the doctors. These are the basic insights gained.

After performing statistical analysis and ml models to get in-depth insights, created a target/prediction variable and performed predictive analysis which gave a good result.

Feature importance analysis Is done using random forest to predict top useful features.

**Conclusion:**

The consumers did not have the full knowledge of the e-vehicles, majority of the consumers are not interested in buying/owning a e-vehicle. The vehicle owned consumers are significantly lower. But seeing the feature importance visualization, the attitude of consumers shows that most of the consumers are having a positive opinion towards e-vehicle performance and think that owning a e-vehicle would lead to cost savings in long run, and are willing to support government new e-vehicle policies. But majority of consumers don’t have the practice with the e-vehicles i.e. owning/renting, identifying a e-vehicle or charging station or research about e-vehicle online.

**improve upon the Market Segmentation Project**

* Collecting more data through conducting surveys, more data leads to better predictions and accuracy.
* More occupations can be added in the occupation feature.
* Creation of new columns through the survey (such as review/experience of e-vehicle owners).
* Data collection over different regions i.e. rural and urban areas, this gives a more information about consumer knowledge and towards e-vehicles.

**Additional ML Models:**

* Unsupervised learning models can be implemented and tested.
* Dimensionality reduction such as T-Sne/PCA to remove the unnecessary features.
* GMM (Gaussian Mixture Model) can be used for better clustering.