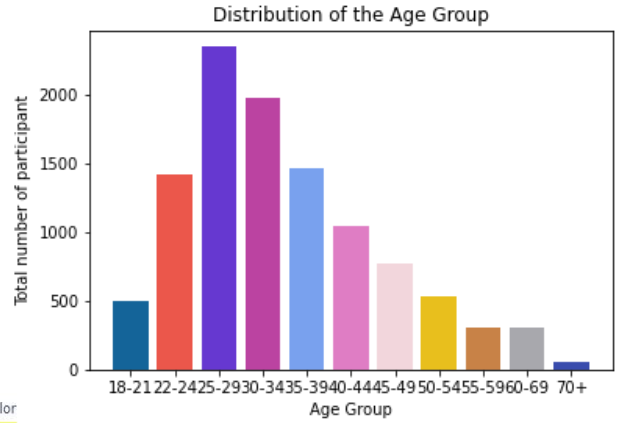
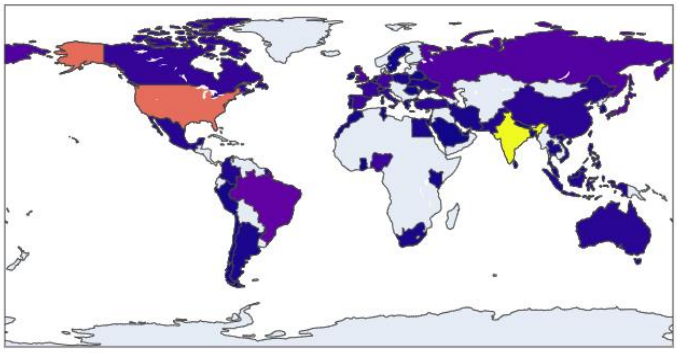
Exploratory Data Analysis:

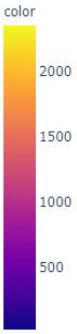
The EDA is performed to analyse the given survey data. For the analysis, Age, Gender, Education, Country, and Salary factors are considered of which the plots are shown below.

First, the distribution of people according to the Age group is generated and the bar plot is used for this purpose. This helps to know what age group has participated most.

It is observed from the plot that most of the participants belongs to the Age group of 25-29.

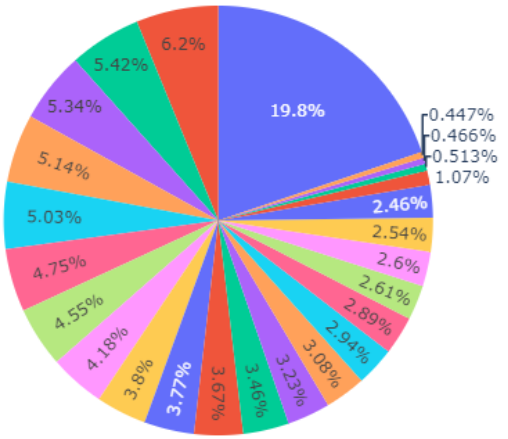


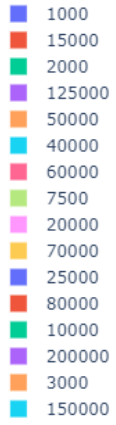




Next, the observation is made for the country where the survey participants belong. For that, the choropleth plot is used, and the country map is shown below.

It is observed that most of the participants are from India and then the USA comes on the second in the list.



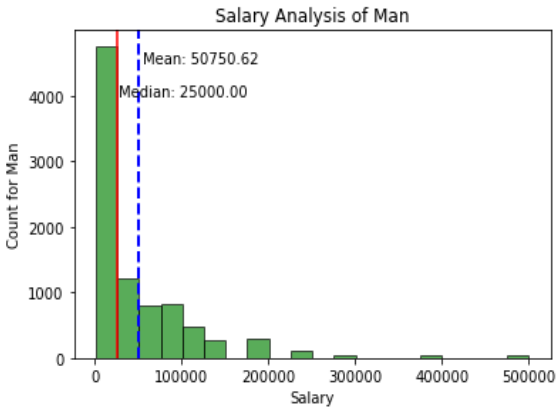


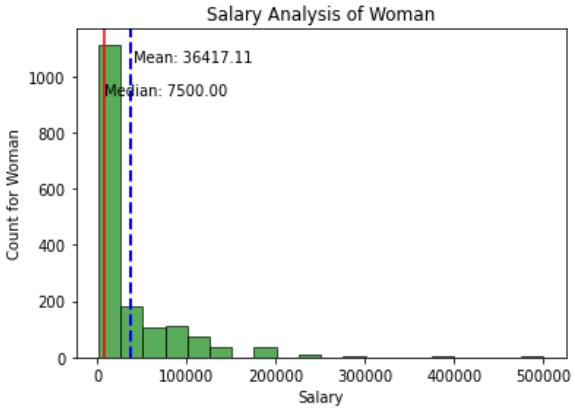
Further for the analysis, the plots for the salary data is generated. For that, initially, the scatter plot is developed initially to look for the salary concentration and then pie plot is developed to know what the maximum salary range of most of the participants is.

The histogram is plotted for that purpose also and noted that most of the salary concentrated in the range of less than $100000. Also, from the pie chart, it is noted that the 19.8% of the people who surveyed has salary of around $1000. Moreover, less than 0.5% people have salary range of $500000. Further, the salary is analysed according to the Age group as shown in the Distribution of Age and Salary scatter plot.

It is noted that mostly the salary of around 400k-500k is earned by the age group of 40-44 because of the experience they may have obtained. Though there are people in other age groups with that amount of salary. Also, the age group of 18-21 are mostly in very low range of salary. This is same with the age group above 70 because of the retirement maybe.

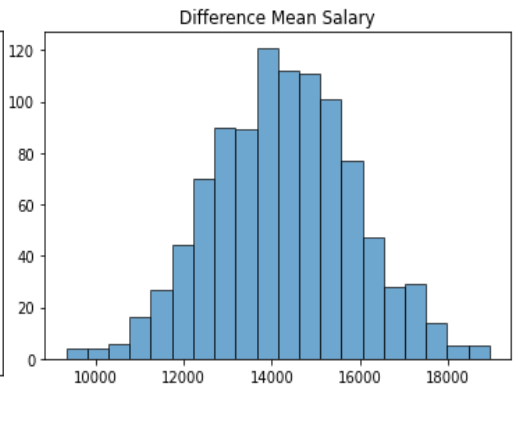
Q2- Answer





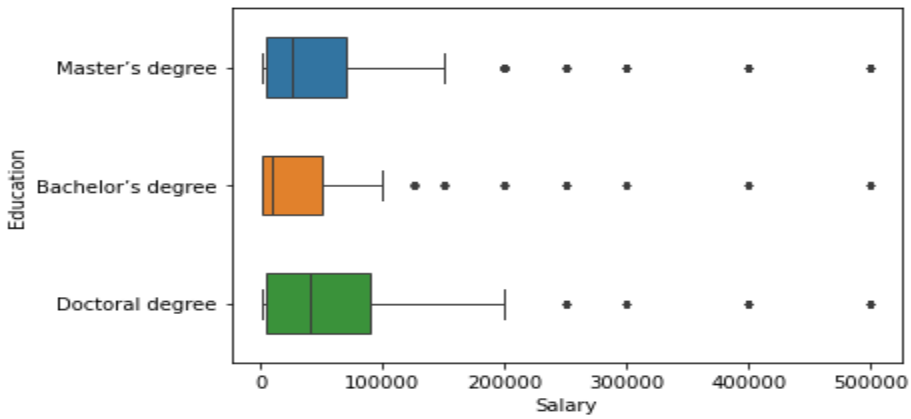
The histogram is plotted for the group of Man and Woman salary and it is found that both the plot is skewed left. Moreover, maximum number of Man and Woman have salary of about $1000. Furthermore, the mean of salary of Man is $50750.62 while that for Woman is $36417.11.

For the hypothesis testing the two-sample t test is performed and found the p value less than 0.05 which indicates that the null hypothesis is rejected, i.e. the mean of two groups is statistically different.



The bootstrap method is performed on both the samples and the histogram plot is generated which shows the same comparison of the mean. It is observed that the Woman generally earns lesser (around 14000 in average) than the Man in Machine learning field. Though there are lesser participance of Woman in this survey, bootstrapping shows the similar trend. Also, for hypothesis testing, the p=0 < 0.05 that implies the mean of two group is statistically different.

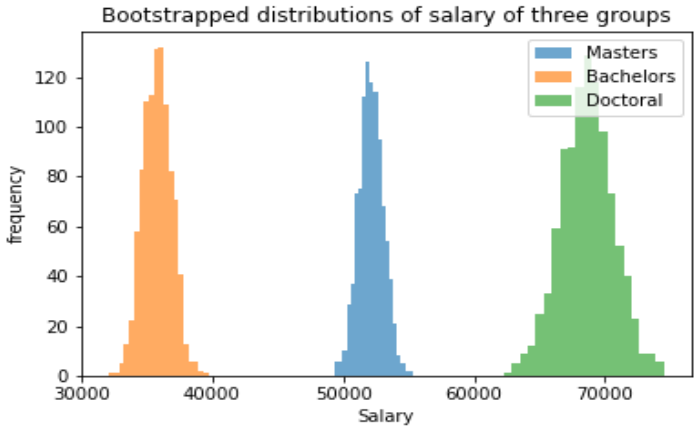
Q3- Answer



The same procedure is adopted as question 2. The statistical distribution and analysis are performed for the three-education group using the box plot.

It is observed that the mean of salary for those who have Doctoral degree is higher than those who have other two degree and the mean of salary of those who have Bachelor's degree is the lowest.

The bootstrapping is performed in the same way as in previous question. Here also, it is noted that the mean of salary for Doctoral degree is the highest and for Bachelor's degree is the lowest. This implies that the education level has impact of the range of salary. Higher the level of education, higher the probability of earning high salary.



The analysis of variance (ANOVA) is performed to compare the mean of three groups. The p value is found to be 0 < 0.05 threshold. Therefore, the hypothesis that the mean of all the three groups is same is rejected.