Note: Before attempting this section, you should have finished part A of the project, which was supplied by Polly.

Part B - Data Visualisation

You are tasked with doing some exploratory data analysis, which is the first step in building a model to predict churn. Since this process is usually very large, we will look at a subset of the total plots you would need to complete this.

1. First you should look at the differences in churn rates, split by the different categorical variables. Produce the appropriate visualisation to compare the average churn rate, split by:
i. Geography
ii. Gender
iii. Tenure
For maximum marks, make sure plots are correctly labelled.

2. We would also like to know how the data is distributed. Some models require features to be normally distributed, and highly skewed variables can affect summary statistics if left unchecked. Produce the appropriate visualisation for the distribution of:
i. Geography
ii. Age
iii. Credit Score

3. Combine all of the above visualisations into a subplot (hint: You may need to use graph objects to recreate some visualisations). For maximum marks, make sure that you correctly label each figure in the subplot.

4. You can get the correlation between all columns using df.corr(). Create a bar plot that shows the correlation of each feature with the target. (Make sure to add a title and axis labels)

4.1. Order the bars so that the feature with the highest correlation is the first bar.

4.2. Add the correlation value to the top of each bar

4.3. Add a line to the figure which shows the average correlation (hint: This will require adding an extra trace).