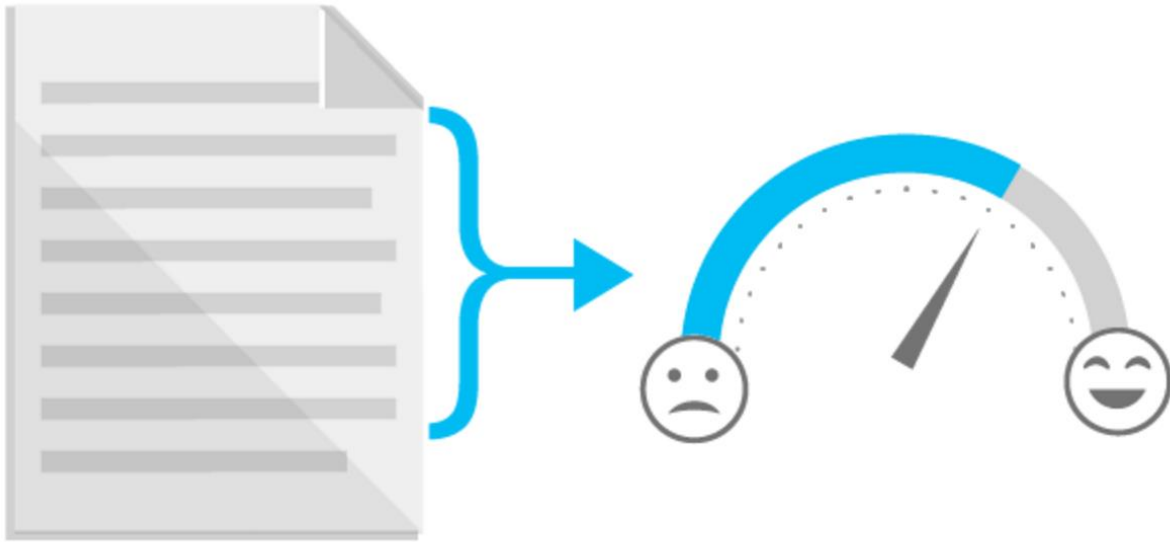


DATA ANALYTICS ASSIGNMENT -2

MICROSOFT AZURE – TEXT ANALYTICS

CED15I030

ESD15I009



About:

The Text Analytics API is a cloud-based service that provides advanced natural language processing over raw text, and includes sentiment analysis. The API is backed by resources in Microsoft cognitive services, a collection of machine learning and AI algorithms in the cloud, readily consumable in our development projects.

X-axis and Y-axis:

X-axis: Review Id

Y-axis: Percentage of positivity

Review Datasets:

Dataset of App: App_Reviews_50.csv

Dataset of Apple products: Apple_Reviews_50.csv

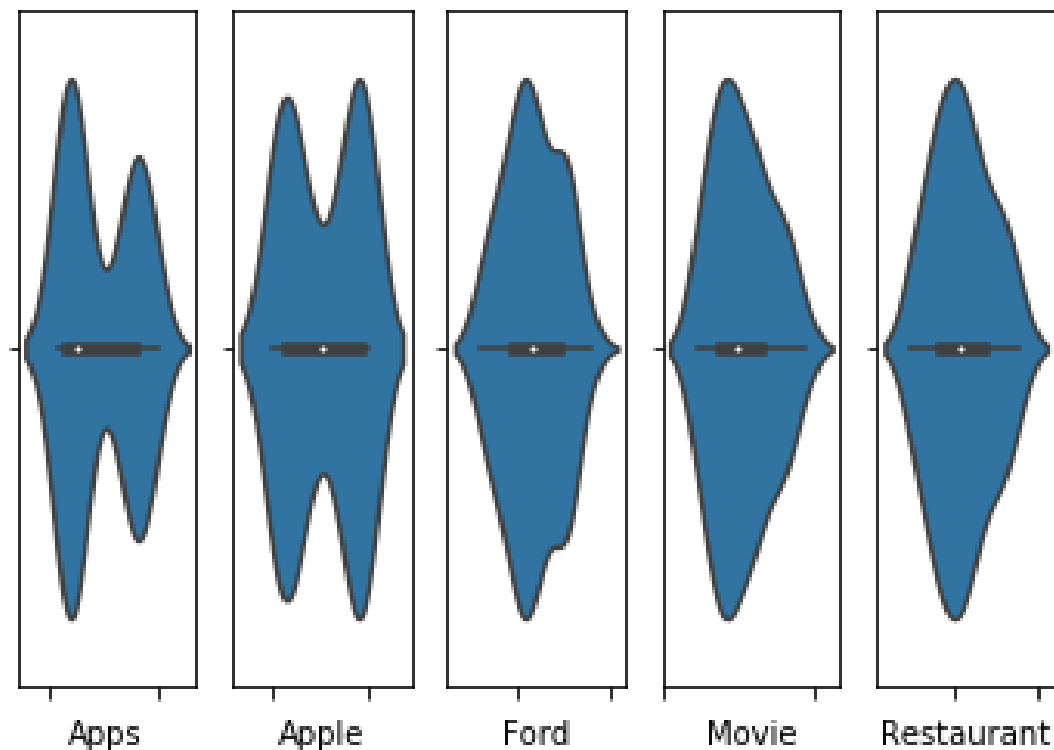
Dataset of Ford Company: Ford_Reviews_50.csv

Dataset of Hollywood movie: Movie_Reviews_50.csv

Dataset of Restaurants: Restaurant_Reviews_50.csv

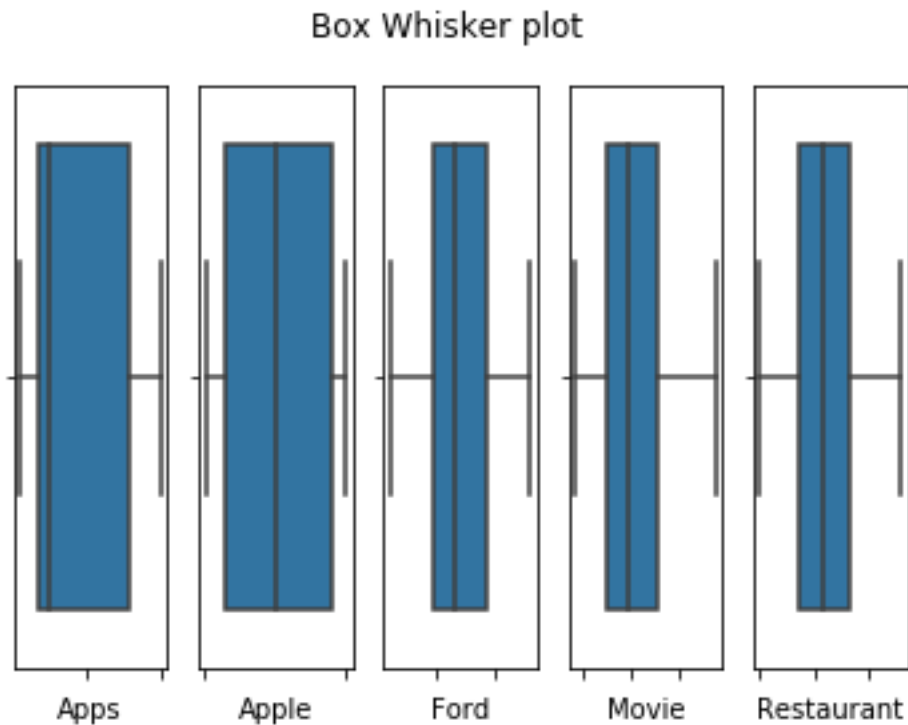
1. **Violin plot:** A **violin plot** is a method of plotting numeric data. It is similar to a box plot with a rotated kernel density plot on each side. A violin plot has four layers. The outer shape represents all possible results, with thickness indicating how common. (Thus the thickest section represents the mode average.). The next layer inside represents the values that occur 95% of the time. The next layer (if it exists) inside represents the values that occur 50% of the time. The central dot represents the median average value.

Violin plot

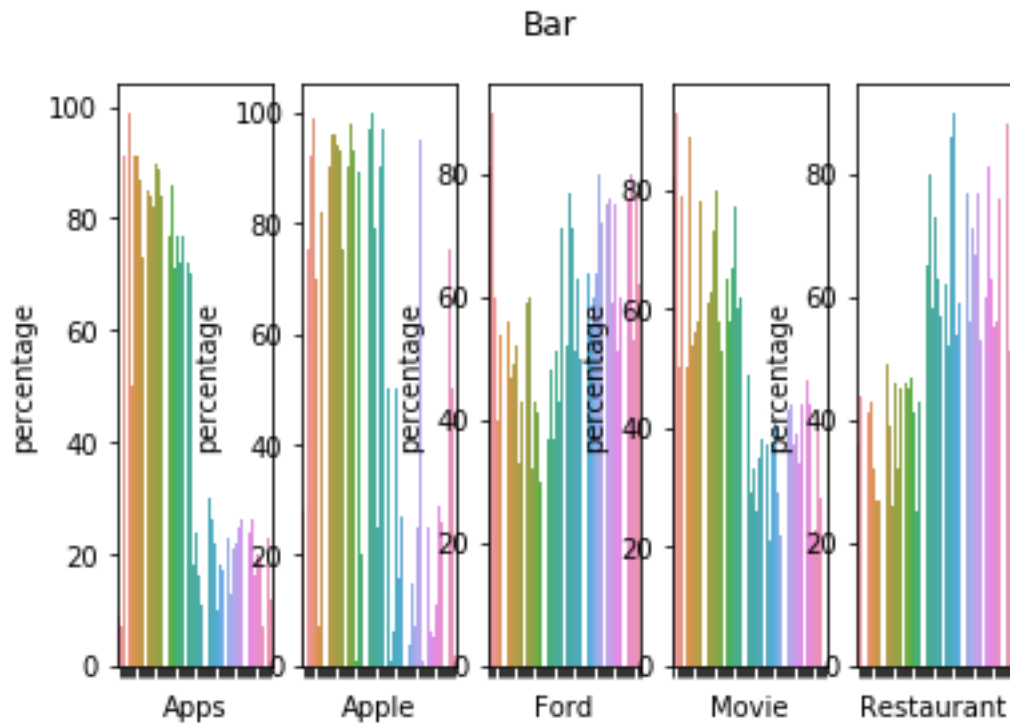


2. **Box – Whisker plot:**

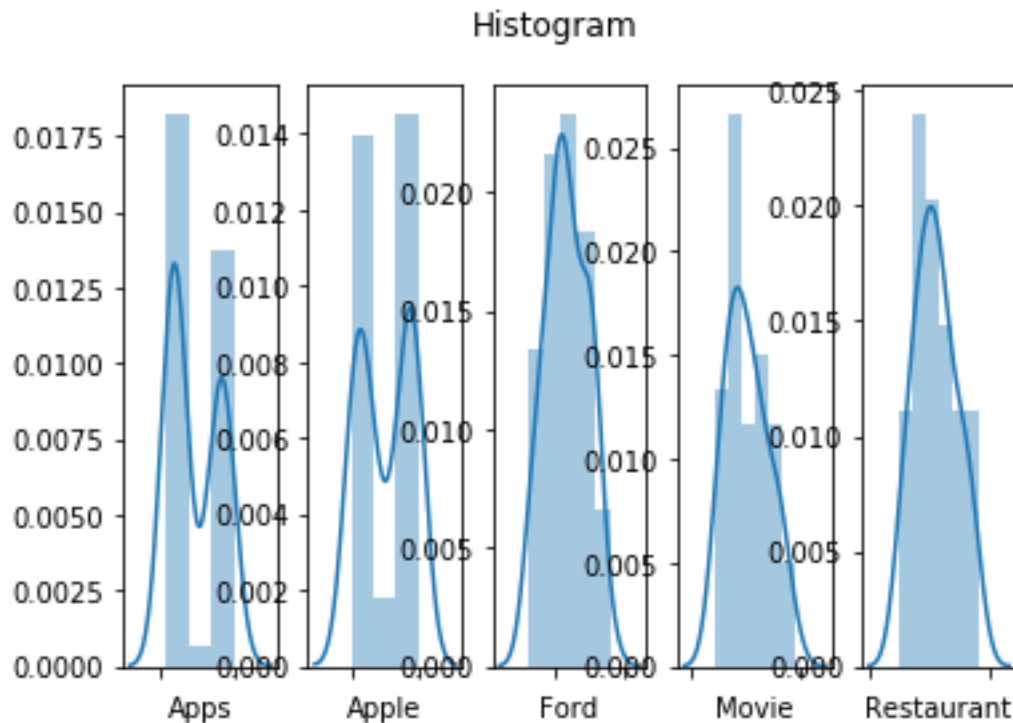
In descriptive statistics, a box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes (*whiskers*) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points. Box plots are non-parametric: they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution (though Tukey's boxplot assumes symmetry for the whiskers and normality for their length).



3. **Bar plot:** A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a line graph.



4. **Histogram:** A histogram is an accurate representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson.^[4] It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one. To construct a histogram, the first step is to "bin" (or "bucket") the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size.



5. **Scatter plot:** A scatter plot (also called a scatterplot, scatter graph, scatter chart, scatter gram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are color-coded, one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

