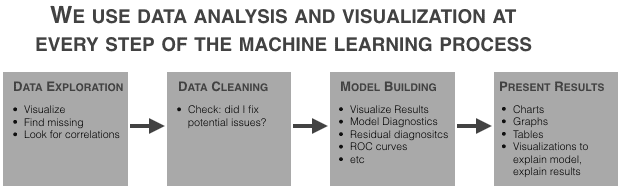
This dataset consists of the OP Waiting List by Group of Hospitals from 2014 - 2018.

The OP Waiting List report shows the total number of patients waiting for a first appointment at a consultant-led Outpatient clinic.



Below are the libraries that are used in order to perform EDA.

**# Importing required libraries.**import pandas as pd  
import numpy as np  
import seaborn as sns #visualisation  
import matplotlib.pyplot as plt #visualisation  
%matplotlib inline   
sns.set(color\_codes=True)

Loading the data into the pandas data frame is certainly one of the most important steps in EDA.The dataset is in CSV format.

To take a closer look at the data took help of “. head ()” function of pandas library which returns first five observations of the data.

there is total 5 observations and 2 columns in the given dataset. Each column represents a variable in the Data Frame. We can see from the data type of each column what type of variable it is. Data cleaning is not required for this data set.

Year Count

count 5.000000 5.000000e+00

mean 2016.000000 5.195765e+06

std 1.581139 7.616284e+05

min 2014.000000 4.236176e+06

25% 2015.000000 4.794370e+06

50% 2016.000000 5.037669e+06

75% 2017.000000 5.782699e+06

max 2018.000000 6.127911e+06

# Column Non-Null Count Dtype

--- ------ -------------- -----

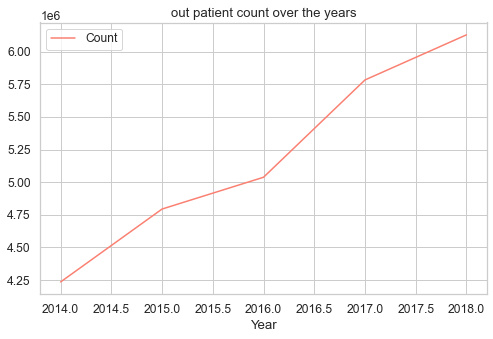
0 Year 5 non-null int64

1 Count 5 non-null int64

dtypes: int64(2)

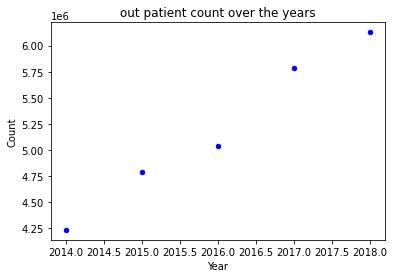
* No Null values present in the given data.
* Data has only float and integer values.
* Now explore data with graphs. Python has a visualization library, Seaborn which build on top of matplotlib. It provides very attractive statistical graphs in order to perform both Univariate and [Multivariate analysis](http://www.camo.com/multivariate_analysis.html).
* data visualisation gives more insights about the data.

**VISUALISATION AND EDA:**



* Here we can clearly see the increase in the count of patients over the years.
* From 2014 to 2015 total no of OP are increased by 4.25 to 4.8 nearly 0.5Million and all the years from 2014 to 2018 4.25 to 6.1Million are in a waiting list for different kind of treatments.

**Scatter plot:**



**Box plot:** The box plot (a.k.a. box and whisker diagram) is a standardized way of displaying the distribution of data based on the five number summary:

Minimum

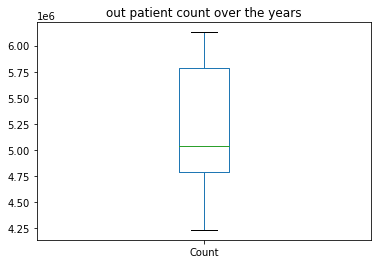
First quartile

Median

Third quartile

Maximum.

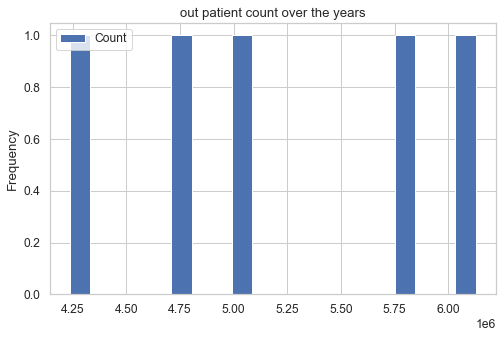
Outliers are either 3×IQR or more above the third quartile or 3×IQR or more below the first quartile.



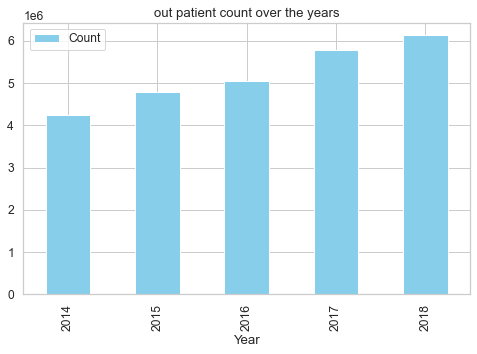
*The boxplot shows the range, the minimum and maximum values with the whiskers, the interquartile range with the box edges, and the median inside.*

* the **mean is greater than** the **median**, the distribution is positively skewed.

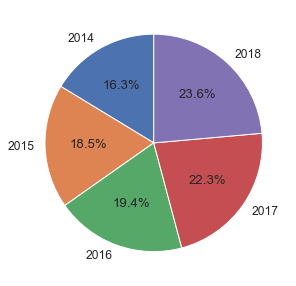
**Histogram**



**Bar Graph**



**Pie Chart**



**Above Pie chart gives the clear idea about the increase in OP patient waiting list year wise in Percentages.**

**From 2014 to 2018 in different hospitals the count of OP waiting list is increased from 16.3% to 23.6%.**