DATA ANALYTICS WITH COGNOS GROUP 2

Phase 3 Submission Document

Project title : Website Traffic Analysis

Name : JAYANTHI.M

Reg No : 420421104030

College: APEC

Project : 3



**Abstract:**

Website traffic analysis is a critical aspect of understanding and optimizing the online presence of businesses and organizations. It involves the collection and examination of data pertaining to user visits, interactions, and engagement with a website. This process encompasses various metrics, such as page views, unique visitors, bounce rates, and conversion rates, which are instrumental in assessing the effectiveness of a website.

**Introduction :**

In the fast-paced digital landscape of the 21st century, a website is often the first point of contact between an organization and its target audience. It serves as a virtual storefront, information hub, and interaction platform, making it crucial for businesses, individuals, and institutions to understand how their websites are performing and how users are engaging with their online content. This understanding is achieved through the process of website traffic analysis.

Website traffic analysis is the systematic examination of data associated with user visits and interactions on a website. It involves collecting, processing, and interpreting a multitude of metrics and data points to gain valuable insights into the effectiveness of a website. This information is essential for making informed decisions about content creation, user experience enhancement, digital marketing strategies, and overall online presence improvement.

**Data visualization:**

Website data analysis visualization is the process of representing and interpreting website-related data through visual elements like charts, graphs, and dashboards. This visual representation of data helps website administrators and analysts gain a clearer understanding of user behavior, website performance, and other critical metrics.

**Necessary step to follow:**

1.Import Libraries:

Start by importing the necessary libraries.

**Program:**

**import** pandas **as** pd

**import** numpy **as** np

**import** seaborn **as** sns

**import** matplotlib.pyplot **as** plt

**from** sklearn.metrics **import** accuracy\_score

**%matplotlib** inline

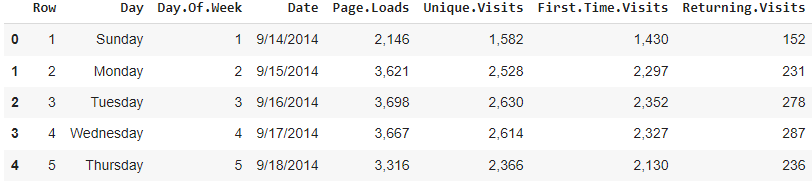
2.Load the Dataset:

Load your into a Pandas DataFrame.You can typically find house price datasets in csv format,but you can adapt this code to other formats as needed.

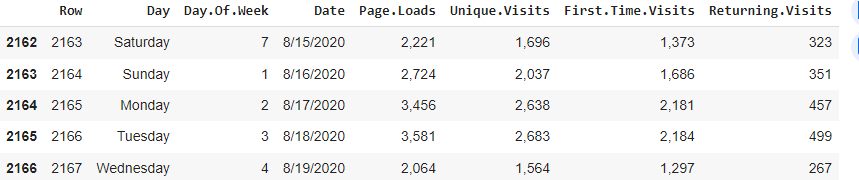
**Program:**

df = pd.read\_csv('daily-website-visitors.csv')

df.head(5)

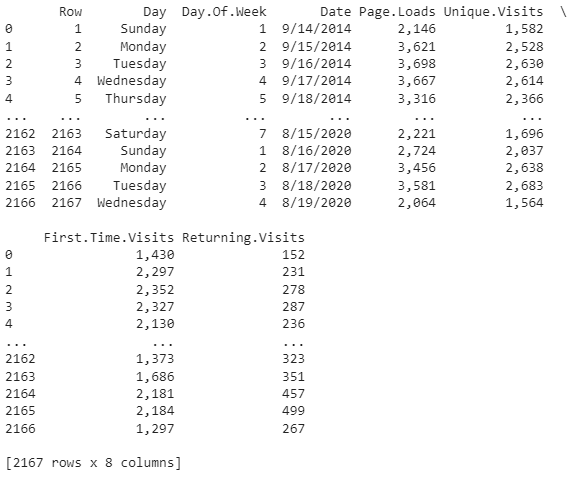
****

df.tail(5)



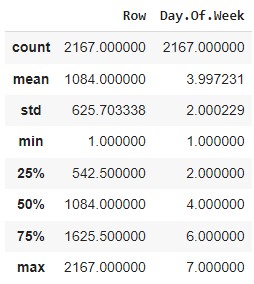
**Take a look at Data:**

print(df)

****

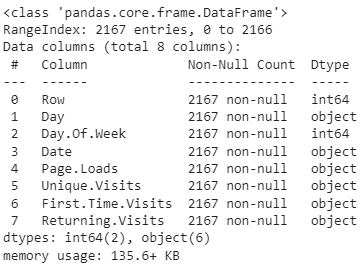
#### Descriptive Statistics:

df.describe()

****

**Dig some information on features, rows and columns:**

df**.**info()

****

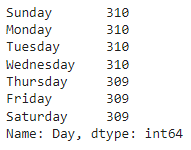
df**.**keys()

****

df**.**shape

****

df['Day'].value\_counts()

****

**Let's Visualize daily website visiters data and get some insight on our dataset!**

Python has lots of visualization tools from Pandas, Matplotlib and Seaborn Libraries. We'll use couples of them to visualize data. You'll find data visualization quite interesting and v interesting fact is without calculating/applying any algorthim we will be able to make intuition on Data.

**Relationship between Day.Of.Week and Page.Loads using scatter plot:**

plt.figure(figsize=(14,8))

ax = df[df.Day=='Sunday'].plot.scatter(x='Day.Of.Week', y='Page.Loads', label='Sunday')

df[df.Day=='Monday'].plot.scatter(x='Day.Of.Week', y='Page.Loads', color='red', label='Monday', ax=ax)

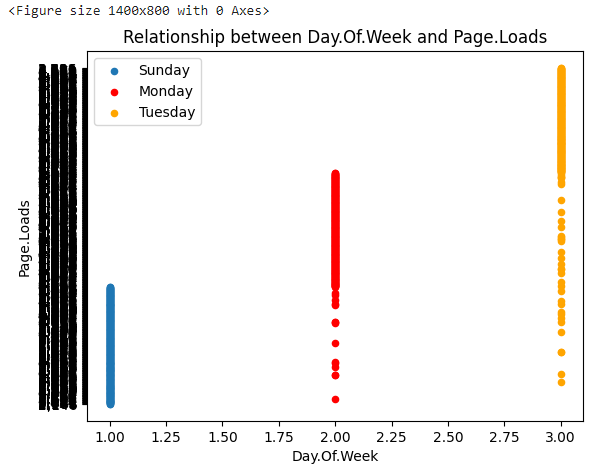
df[df.Day=='Tuesday'].plot.scatter(x='Day.Of.Week', y='Page.Loads', color='orange', label='Tuesday', ax=ax)

ax.set\_xlabel("Day.Of.Week")

ax.set\_ylabel("Page.Loads")

ax.set\_title("Relationship between Day.Of.Week and Page.Loads")

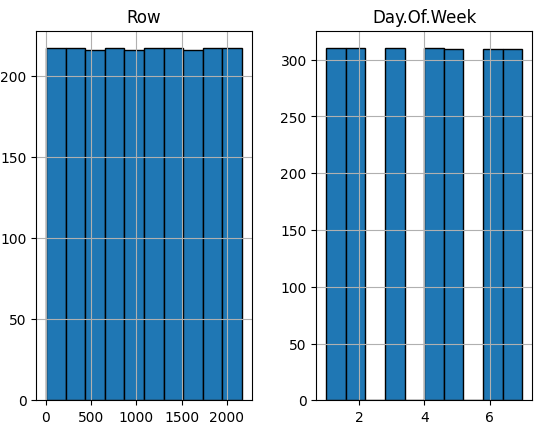
plt.show()

****

**Histogram:**

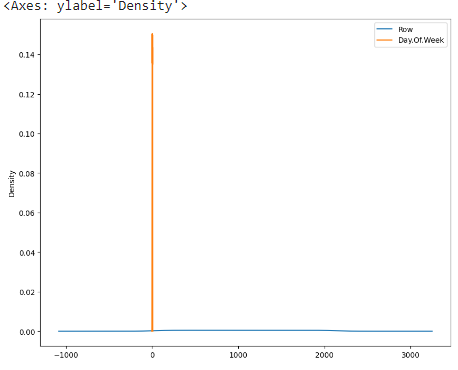
df**.**hist(edgecolor**=**'black', linewidth**=**1)

plt**.**show()

****

#### Take a look at density:

df**.**plot(kind **=** "density", figsize**=**(10,8))

**

### Seaborn:

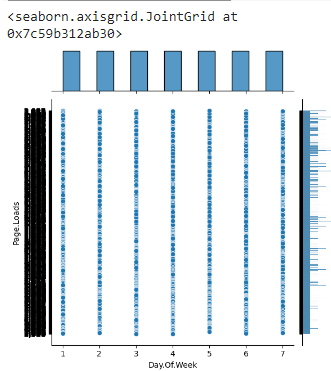
Seaborn has many nice functions for plotting. You can create jointplot, hisplot, pairplot, boxplot, violinplot(it's damn cool), heatmap polt etc etc. I recommend seaborn personally cause it has many cool ploting techniques!

**Program:**

import seaborn as sns

df = pd.read\_csv('daily-website-visitors.csv')

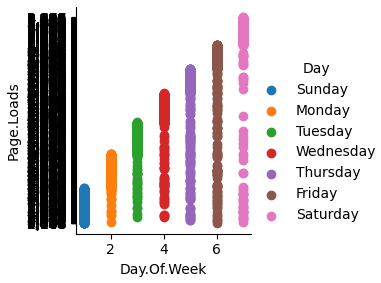
sns.jointplot(x='Day.Of.Week',y='Page.Loads', data = df)



#Seaborn **Scatterplot** by Day onDay.Of.Week vs Page.Loads

g = sns.FacetGrid(df, hue='Day')

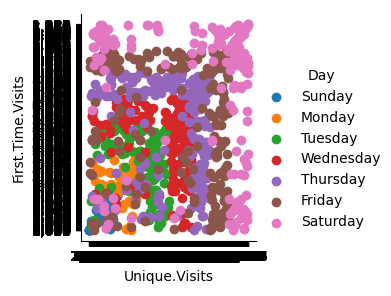
g = g.map(plt.scatter, 'Day.Of.Week','Page.Loads').add\_legend()



# seaborn scatterplot by Day onUnique.Visits vs First.Time.Visits

g = sns.FacetGrid(df, hue='Day')

g = g.map(plt.scatter, 'Unique.Visits', 'First.Time.Visits').add\_legend()

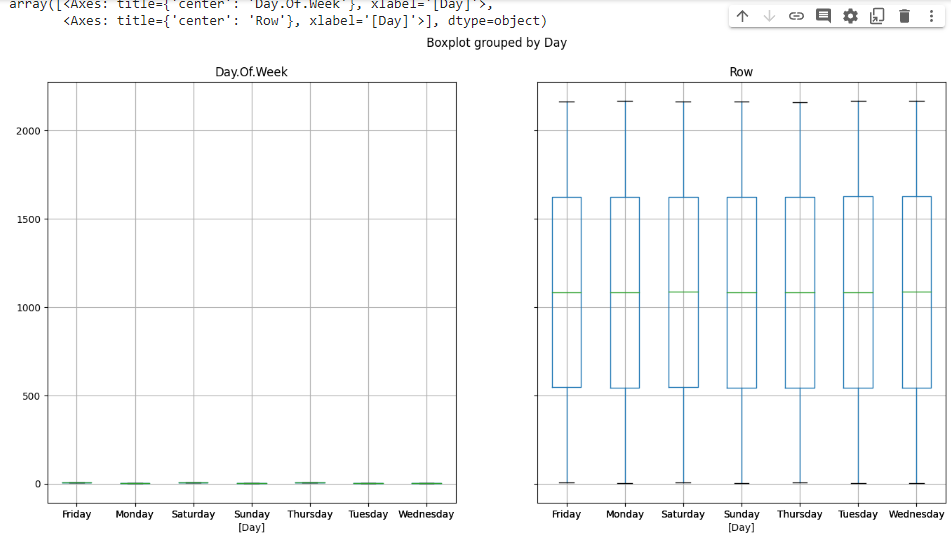
****

**Box Plot:**

***Box Plot*** is the visual representation of the depicting groups of numerical data through their quartiles. Boxplot is also used for detect the outlier in data set. It captures the summary of the data efficiently with a simple box and whiskers and allows us to compare easily across groups.

**Program:**

df.boxplot(by = 'Day', figsize = (16,8))



**Voilin Plot:**

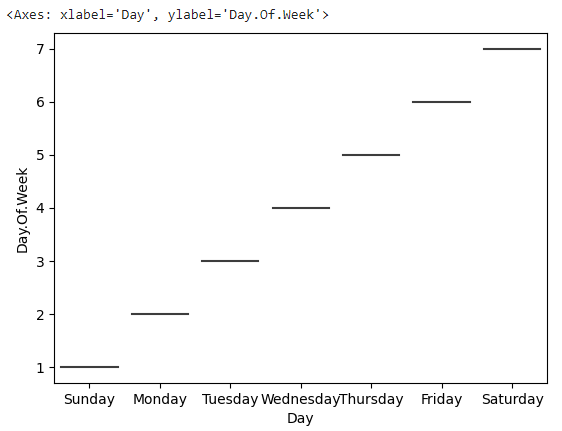
A violin plot plays a similar role as a box-and-whisker plot. It shows the distribution of data points after grouping by one (or more) variables. Unlike a box plot, each violin is drawn using a kernel density estimate of the underlying distribution.

**Program:**

plt.figure(figsize=(14,10))

plt.subplot(2,2,1)

sns.violinplot(x='Day', y='Day.Of.Week', data=df, size=5)



**Conclusion:**

This paper described various measurement tools and modeling techniques for evaluating Web server performance. Web Log Analyzer tools are a part of Web Analytics Software. They take a log file as an input, analyze it and generate results. A variety of tools are available which offer great capabilities in reporting the results of analysis. A study was done and some of the tools were studied. Every tool offered some or the other feature which was better than the rest.