DAC PHASE 4

Website Traffic Analysis

**Introduction:**

Website traffic analysis is a critical component of understanding the performance and effectiveness of a website. In the digital age, websites serve as the primary interface between businesses, organizations, and their target audience. Analyzing website traffic provides invaluable insights into user behavior, preferences, and the impact of marketing efforts. This process involves the collection and examination of data related to the number of visitors, their demographics, sources of traffic, and their interactions with the site.

**First Time Visits:**

* The number of first time visits is increasing over time. This is a good sign that your site is attracting new visitors.
* The highest number of first time visits occurs on Sundays and Mondays. This may suggest that people are more likely to visit your site on the weekends and at the beginning of the week.

**Page Loads:**

* The number of page loads is highest on Sundays and Mondays, followed by Saturdays. This is consistent with the trend observed for first time visits.
* The number of page loads is lowest on Wednesdays and Thursdays.

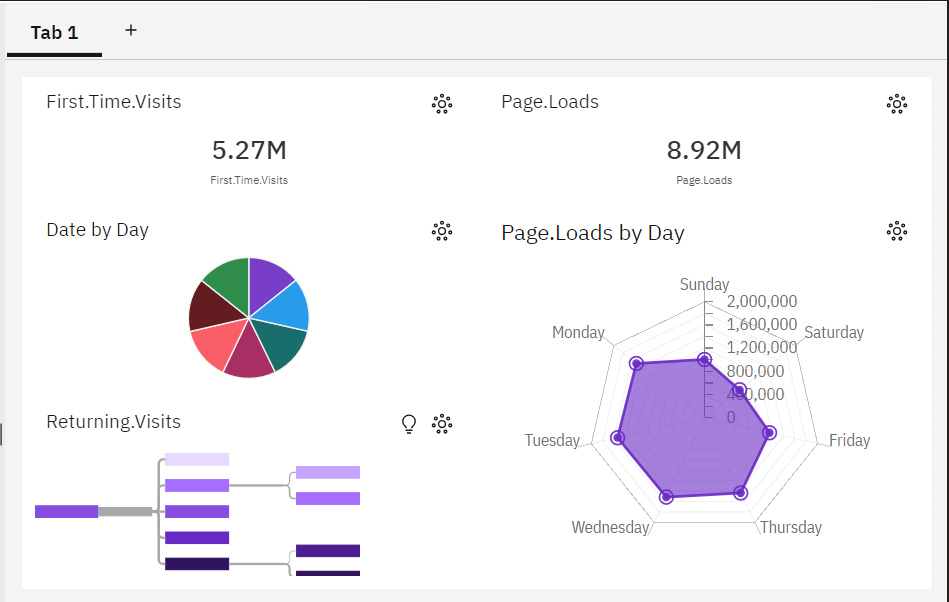
**Returning Visits:**

* The number of returning visits is increasing over time. This is another good sign that your site is engaging visitors and encouraging them to come back.
* The highest number of returning visits occurs on Sundays and Mondays. This is consistent with the trend observed for first time visits and page loads.

**Additional insights:**

* The ratio of first time visits to page loads is decreasing over time. This suggests that visitors are becoming more likely to visit multiple pages per visit.
* The ratio of returning visits to page loads is increasing over time. This suggests that returning visitors are becoming more engaged with your site.

**Visualization:**



Python Code:

import numpy as np # linear algebra  
import pandas as pd   
import os  
for dirname, \_, filenames in os.walk('/content/daily-website-visitors.csv

):  
 for filename in filenames:  
 print(os.path.join(dirname, filename))

from sklearn.preprocessing import MinMaxScaler, StandardScaler, PolynomialFeatures  
from sklearn.model\_selection import train\_test\_split  
from sklearn.metrics import mean\_absolute\_error as mae, mean\_squared\_error as mse, mean\_squared\_log\_error as msle, r2\_score  
from sklearn.linear\_model import LinearRegression, Ridge, Lasso  
from sklearn.ensemble import RandomForestRegressor  
from sklearn.tree import DecisionTreeRegressor  
from xgboost import XGBRegressor

# The Qustion is to predict 'Returning.Visits'

# 

# 1. Check dataset

data = pd.read\_csv('/content/daily-website-visitors.csv

)  
data.head()

Row Day Day.Of.Week Date Page.Loads Unique.Visits \  
0 1 Sunday 1 9/14/2014 2,146 1,582   
1 2 Monday 2 9/15/2014 3,621 2,528   
2 3 Tuesday 3 9/16/2014 3,698 2,630   
3 4 Wednesday 4 9/17/2014 3,667 2,614   
4 5 Thursday 5 9/18/2014 3,316 2,366   
  
 First.Time.Visits Returning.Visits   
0 1,430 152   
1 2,297 231   
2 2,352 278   
3 2,327 287   
4 2,130 236

# 

# 2. To change Columns name to use easly (delete '.')

data.columns = data.columns.str.replace('.', '', regex = True)

# 3. To Check data types and missing data

data.info() # There is no missing data

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2167 entries, 0 to 2166  
Data columns (total 8 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Row 2167 non-null int64   
 1 Day 2167 non-null object  
 2 DayOfWeek 2167 non-null int64   
 3 Date 2167 non-null object  
 4 PageLoads 2167 non-null object  
 5 UniqueVisits 2167 non-null object  
 6 FirstTimeVisits 2167 non-null object  
 7 ReturningVisits 2167 non-null object  
dtypes: int64(2), object(6)  
memory usage: 135.6+ KB

data.head(5)

Row Day DayOfWeek Date PageLoads UniqueVisits \  
0 1 Sunday 1 9/14/2014 2,146 1,582   
1 2 Monday 2 9/15/2014 3,621 2,528   
2 3 Tuesday 3 9/16/2014 3,698 2,630   
3 4 Wednesday 4 9/17/2014 3,667 2,614   
4 5 Thursday 5 9/18/2014 3,316 2,366   
  
 FirstTimeVisits ReturningVisits   
0 1,430 152   
1 2,297 231   
2 2,352 278   
3 2,327 287   
4 2,130 236

'Day' Column and 'DayOfWeek' Column look similar, so I would drop 'Day' Column beacuse the type of 'Day' column is 'object'

data = data.drop('Day', axis = 1)

'PageLoads' ~ 'ReturningVisits' columns need to be changed 'int' type

col = ['PageLoads', 'UniqueVisits', 'FirstTimeVisits', 'ReturningVisits']  
  
data[col] = data[col].replace(',', '', regex = True).astype('int64')

'Date' Column needs to be changed 'datetime64'

data['Date'] = data['Date'].astype('datetime64')  
data['year'] = data['Date'].dt.year  
data['month'] = data['Date'].dt.month  
data['day'] = data['Date'].dt.day  
data.drop(columns = 'Date', inplace = True)

data.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2167 entries, 0 to 2166  
Data columns (total 9 columns):  
 # Column Non-Null Count Dtype  
--- ------ -------------- -----  
 0 Row 2167 non-null int64  
 1 DayOfWeek 2167 non-null int64  
 2 PageLoads 2167 non-null int64  
 3 UniqueVisits 2167 non-null int64  
 4 FirstTimeVisits 2167 non-null int64  
 5 ReturningVisits 2167 non-null int64  
 6 year 2167 non-null int64  
 7 month 2167 non-null int64  
 8 day 2167 non-null int64  
dtypes: int64(9)  
memory usage: 152.5 KB

**Conclusion:**

In conclusion, website traffic analysis is a vital tool for harnessing the power of the digital landscape. It allows organizations to adapt and evolve in response to changing user trends and preferences. By comprehensively understanding their audience, businesses can tailor their content and marketing strategies, thereby enhancing user engagement and conversion rates. The data obtained from website traffic analysis can help identify areas for improvement, such as optimizing site speed, improving user interfaces, and refining content strategies. In an ever-evolving online world, the ability to harness the insights derived from website traffic analysis is an indispensable asset for achieving success and maintaining a competitive edge in the digital realm.