**Phase 5: Final Submission**

**Introduction**

Website traffic analysis is the process of monitoring and evaluating the data related to the visitors to a website. This analysis provides valuable insights into how users interact with your website, what content they consume, where they come from, and much more. The behavior pattern can be analyzed using the python.

IBM Cognos is a business intelligence and performance management software suite developed by IBM. It is designed to help organizations make informed business decisions by providing capabilities for reporting, analytics, dashboards, scorecards, and data visualization. IBM Cognos is used by businesses and enterprises to transform data from various sources into valuable insights and actionable information.

**Design Thinking:**

1. Analysis Objectives: The specific dataset consists of website traffic data, such as identifying popular pages, traffic trends, and user engagement metrics.
2. Data Collection: The excel sheet consists of 7 columns with rows day, day of week, date, page load, unique visits, first time visits and returning visits.
3. Visualization: The data in the data can be visualized using IBM Cognos to create meaningful dashboards and reports. The datas which can be visualized are:
4. How many visitors in a particular day?
5. How popular the website is?
6. Haw many new visitors have visited the website?
7. Unique visitors in a particular week

**Libraries used for data analysis:**

**Pandas:** Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

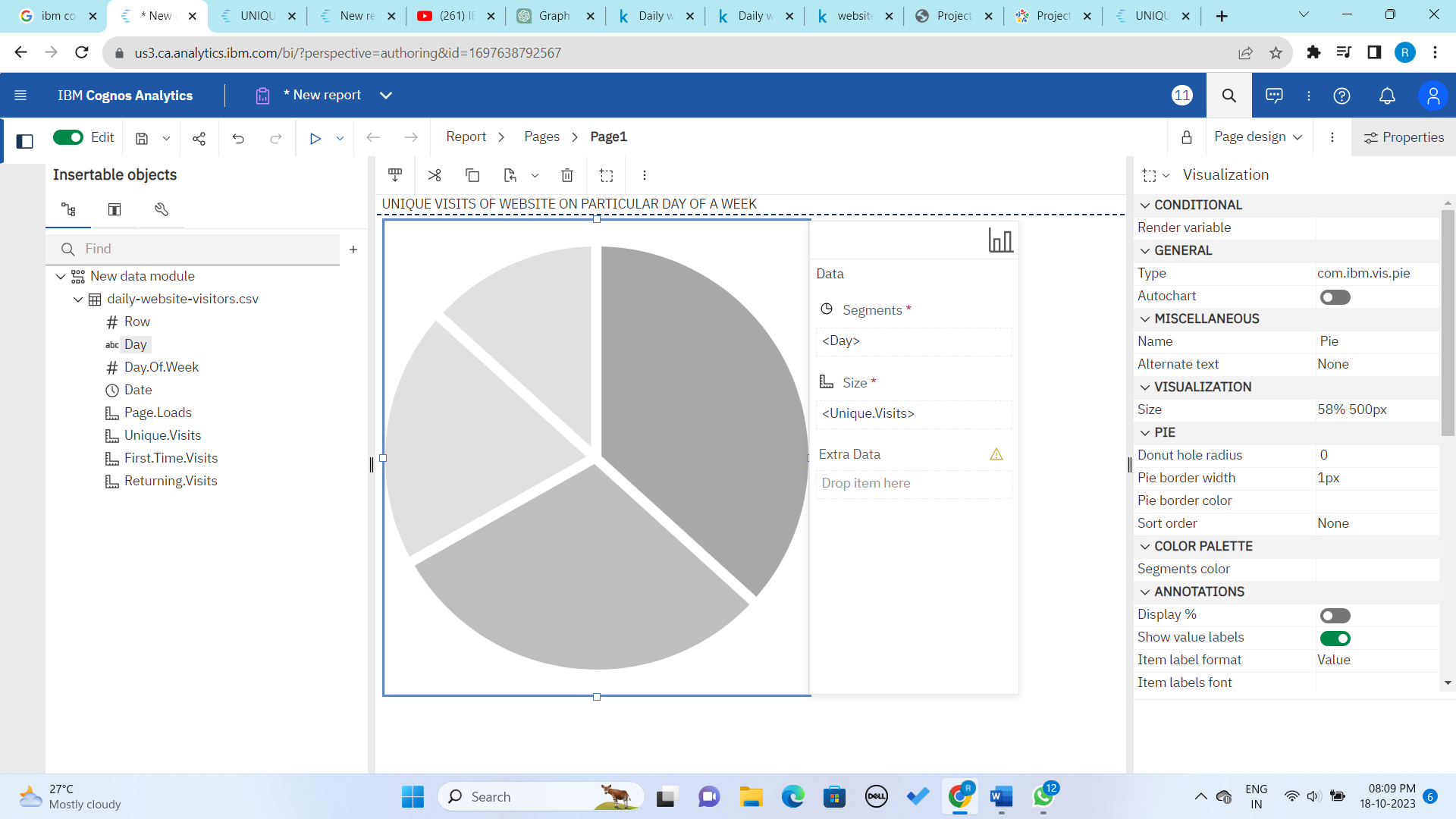
**NumPy:** NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

**Matplotlib Pyplot:** Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

**Seaborn:** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

**Data visualizations using IBM Cognos:**

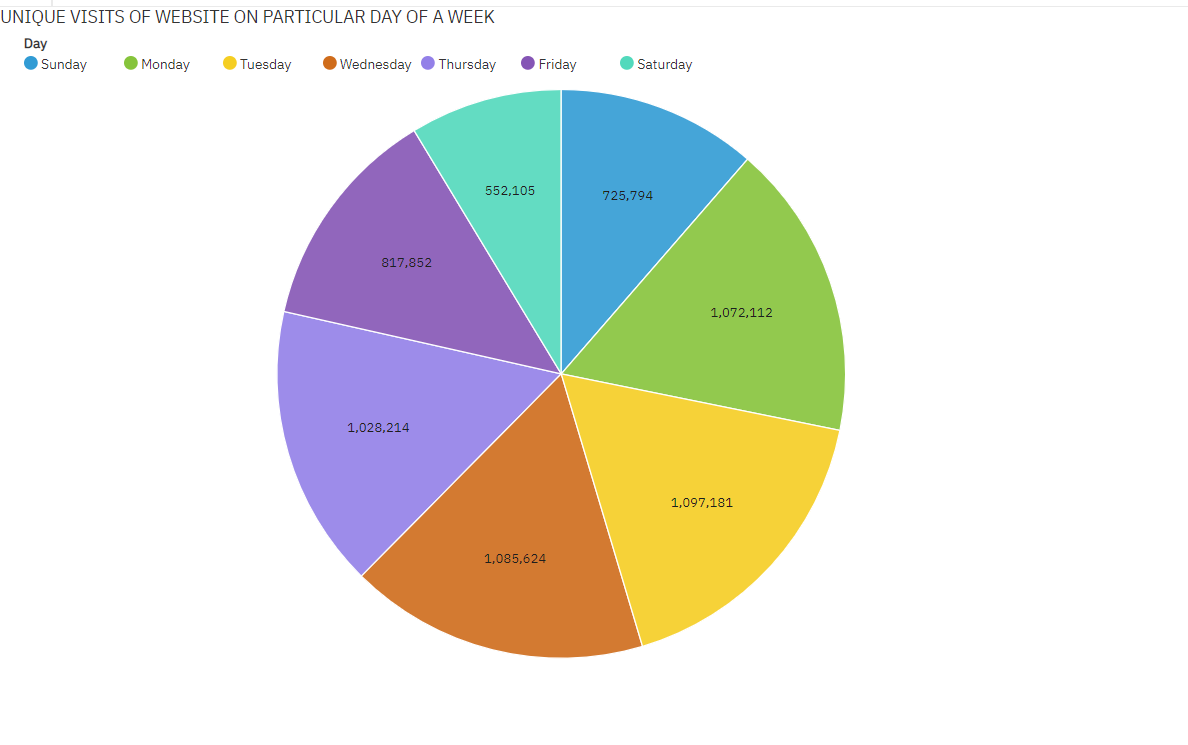
**Report:**



**Output:**

**Link**[**:** https://us3.ca.analytics.ibm.com/bi/?pathRef=.my\_folders%2FUNIQUE%2BVISITS%2BOF%2BWEBSITE%2BON%2BPARTICULAR%2BDAY%2BOF%2BA%2BWEEK&action=run&format=HTML&prompt=false](:%20https:/us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2FUNIQUE%2BVISITS%2BOF%2BWEBSITE%2BON%2BPARTICULAR%2BDAY%2BOF%2BA%2BWEEK&action=run&format=HTML&prompt=false)

**Created report:**

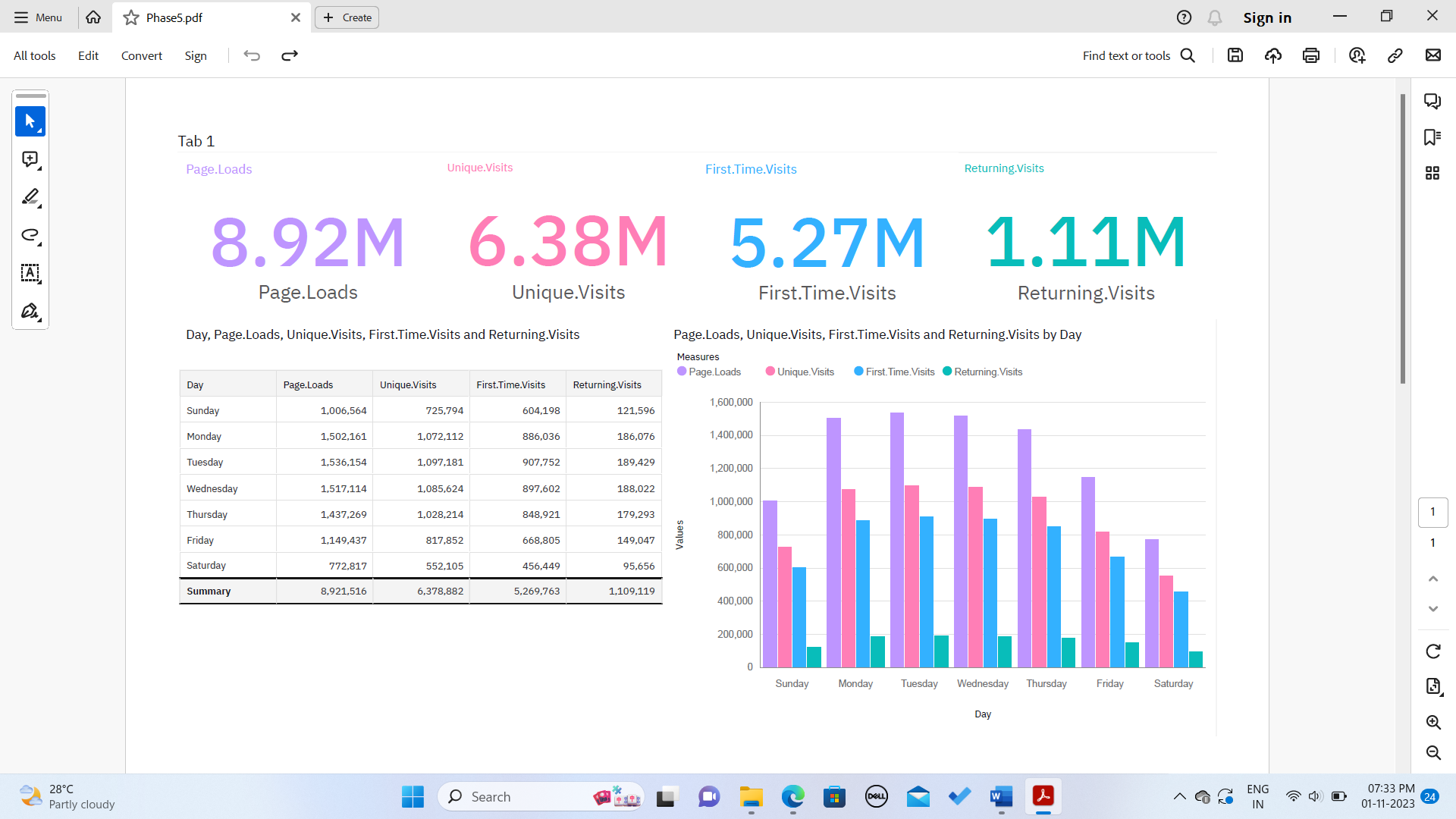


Pie charts are useful when you want to show the composition of your website traffic. From the above pie chart, we can observe that

1. Saturday has low number of unique visits
2. Tuesday has more number of unique visits

**Dashboard**

The following dashboard will show the details about page loads, unique visits, first time visits and returning visits of a website over a week. It will also display the details regarding the total website weightage.



<https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FPhase5&action=view&mode=dashboard&subView=model0000018b8977fa18_00000000>

**Python program:**

The preprocessing of dataset uses libraries such as pandas, numpy etc. The basic preprocessing code of python

1. **Display the dataset**

*import pandas as pd*

*website\_data=pd.read\_csv("D:\Data Analytics\daily-website-visitors.csv")*

*print(website\_data)*

**Output:**

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

... ... ... ... ... ... ...

2162 2163 Saturday 7 8/15/2020 2,221 1,696

2163 2164 Sunday 1 8/16/2020 2,724 2,037

2164 2165 Monday 2 8/17/2020 3,456 2,638

2165 2166 Tuesday 3 8/18/2020 3,581 2,683

2166 2167 Wednesday 4 8/19/2020 2,064 1,564

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

... ... ...

2162 1,373 323

2163 1,686 351

2164 2,181 457

2165 2,184 499

2166 1,297 267

1. **Display specified number of rows**

*import matplotlib.pyplot as plt*

*import pandas as pd*

*import seaborn as sns*

*import numpy as np*

*data=pd.read\_csv("D:\Data Analytics\daily-website-visitors.csv")*

*data.head(7)*

**Output**

Row Day Day.Of.Week Date Page.Loads Unique.Visits First.Time.Visits Returning.Visits

0 1 Sunday 1 9/14/2014 2,146 1,582 1,430 152

1 2 Monday 2 9/15/2014 3,621 2,528 2,297 231

2 3 Tuesday 3 9/16/2014 3,698 2,630 2,352 278

3 4 Wednesday 4 9/17/2014 3,667 2,614 2,327 287

4 5 Thursday 5 9/18/2014 3,316 2,366 2,130 236

5 6 Friday 6 9/19/2014 2,815 1,863 1,622 241

6 7 Saturday 7 9/20/2014 1,658 1,118 985 133

1. **Remove null values**

*data=data.dropna()*

*data.isnull().sum()*

**Output**

Row 0

Day 0

Day.Of.Week 0

Date 0

Page.Loads 0

Unique.Visits 0

First.Time.Visits 0

Returning.Visits 0

dtype: int64

**Bar chart using Python**

*import matplotlib.pyplot as plt*

*import pandas as pd*

*import seaborn as sns*

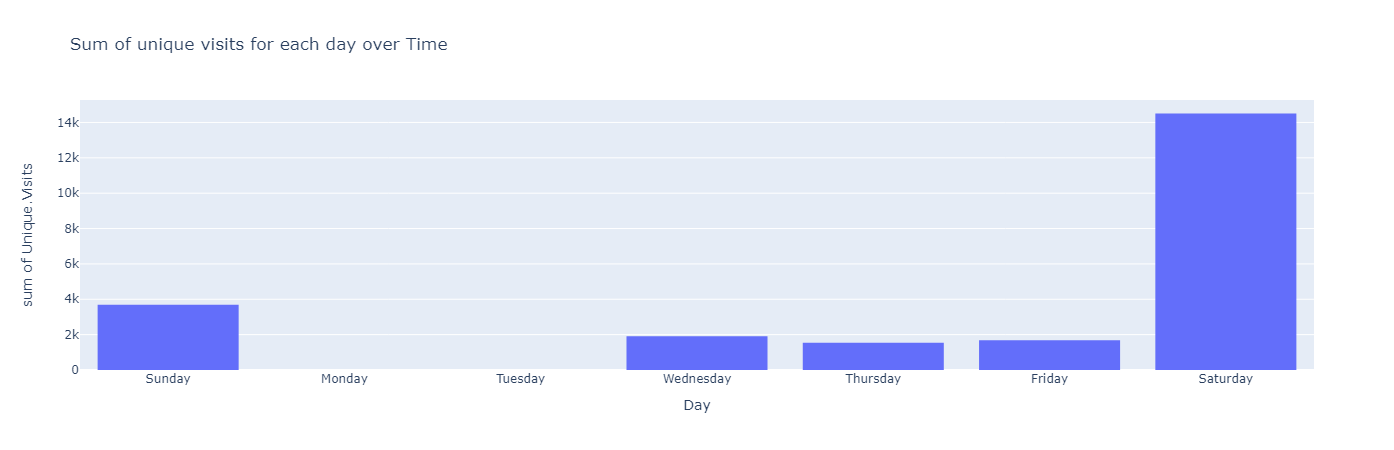
*import numpy as np*

*data=pd.read\_csv("D:\Data Analytics\daily-website-visitors.csv")*

*data.head()*

*import plotly.express as px*

*px.histogram(data,x='Day',y='Unique.Visits',title='Sum of unique visits for each day over Time')*



**Conclusion:**

Thus the data is pre-processed and manipulated to bar chart using python and IBM Cognos