

WEBSITE TRAFFIC ANALYSIS

DATA ANALYTICS
WITH COGNOS



AGENDA

- Introduction
- Problem statements
- IBM visualization
- Python codes
- Conclusion



introduction

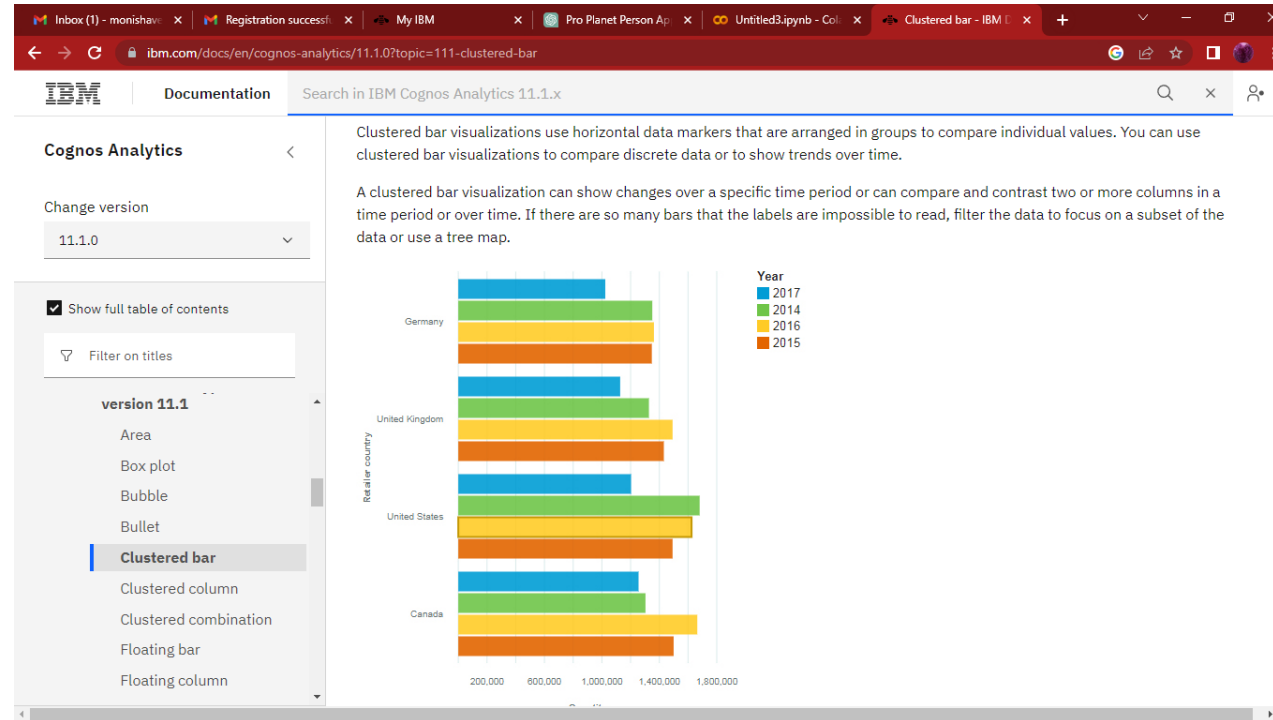
- A website traffic analysis project is an essential component of understanding and optimizing the performance of a website. It involves the systematic collection, measurement, and interpretation of data related to user interactions and behavior on a website. By analyzing website traffic, businesses and website owners can gain valuable insights into their audience, content effectiveness, and overall website performance



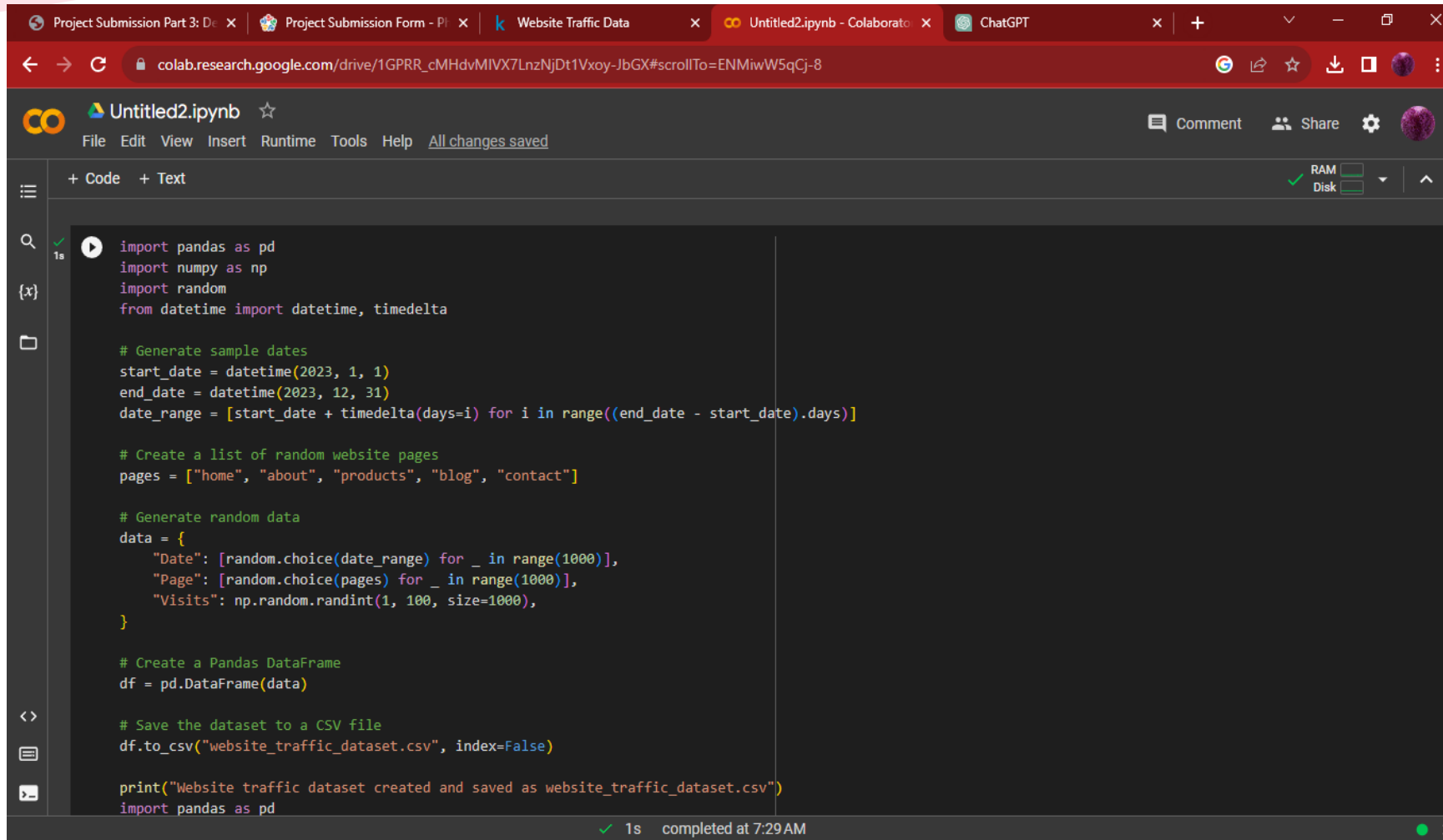
Problem statements

- "Many businesses and website owners face the challenge of understanding and optimizing their website's performance to meet their online goals. They lack the necessary insights into user behavior, content effectiveness, and key performance indicators, making it difficult to make data-driven decisions and improvements. This project aims to address this issue by conducting a comprehensive website traffic analysis to provide actionable recommendations for enhancing the user experience, increasing conversions, and improving overall website performance."

IBM VISUALIZATION



Python codes



The screenshot shows a Google Colab notebook interface. The browser tabs at the top include 'Project Submission Part 3: D...', 'Project Submission Form - P...', 'Website Traffic Data', 'Untitled2.ipynb - Colaborato...', and 'ChatGPT'. The address bar shows the URL 'colab.research.google.com/drive/1GPRR_cMHdvMIVX7LnzNjDt1Vxoy-JbGX#scrollTo=ENMiwW5qCj-8'. The notebook title is 'Untitled2.ipynb'. The menu bar includes 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', 'Help', and 'All changes saved'. The toolbar shows '+ Code' and '+ Text' buttons, along with RAM and Disk usage indicators. The code editor contains the following Python code:

```
import pandas as pd
import numpy as np
import random
from datetime import datetime, timedelta

# Generate sample dates
start_date = datetime(2023, 1, 1)
end_date = datetime(2023, 12, 31)
date_range = [start_date + timedelta(days=i) for i in range((end_date - start_date).days)]

# Create a list of random website pages
pages = ["home", "about", "products", "blog", "contact"]

# Generate random data
data = {
    "Date": [random.choice(date_range) for _ in range(1000)],
    "Page": [random.choice(pages) for _ in range(1000)],
    "Visits": np.random.randint(1, 100, size=1000),
}

# Create a Pandas DataFrame
df = pd.DataFrame(data)

# Save the dataset to a CSV file
df.to_csv("website_traffic_dataset.csv", index=False)

print("Website traffic dataset created and saved as website_traffic_dataset.csv")
import pandas as pd
```

The status bar at the bottom indicates '1s completed at 7:29 AM'.

Project Submission Part 3: De xProject Submission Form - Ph xWebsite Traffic Data xUntitled2.ipynb - Colaborato xChatGPT x

colab.research.google.com/drive/1GPRR_cMHdvMIVX7LnzNjDt1Vxoy-JbGX#scrollTo=-k7e4dJv81fs

Untitled2.ipynb

CommentShare

FileEditViewInsertRuntimeToolsHelpAll changes saved

+ Code + Text

RAMDisk

1s

Generate sample dates

start_date = datetime(2023, 1, 1)

end_date = datetime(2023, 12, 31)

date_range = [start_date + timedelta(days=i) for i in range((end_date - start_date).days)]

Create a list of random website pages

pages = ["home", "about", "products", "blog", "contact"]

Generate random data

data = {

"Date": [random.choice(date_range) for _ in range(1000)],

"Page": [random.choice(pages) for _ in range(1000)],

"Visits": np.random.randint(1, 100, size=1000),

}

Create a Pandas DataFrame

df = pd.DataFrame(data)

Save the dataset to a CSV file

df.to_csv("website_traffic_dataset.csv", index=False)

print("Website traffic dataset created and saved as website_traffic_dataset.csv")

import pandas as pd

import numpy as np

import random

from datetime import datetime, timedelta

Generate sample dates

start_date = datetime(2023, 1, 1)

1s completed at 7:29 AM

Project Submission Part 3: De... xProject Submission Form - Pl... xWebsite Traffic Data xUntitled2.ipynb - Colaborato... xChatGPT x

colab.research.google.com/drive/1GPRR_cMHdvMIVX7LnzNjDt1Vxoy-JbGX#scrollTo=rSaYRo1wAgeW

Untitled2.ipynb ☆
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

✓ 1s

```
df.to_csv("website_traffic_dataset.csv", index=False)

print("Website traffic dataset created and saved as website_traffic_dataset.csv")
```

Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv
Website traffic dataset created and saved as website_traffic_dataset.csv

✓ 0s

```
import pandas as pd

# Load the dataset from the CSV file
df = pd.read_csv("website_traffic_dataset.csv")

# Display the first few rows of the dataset
print(df.head())
```

	Date	Page	Visits
0	2023-08-07	about	39
1	2023-05-31	products	38
2	2023-10-16	products	65
3	2023-12-22	about	60
4	2023-10-27	blog	71

✓ 1s completed at 7:29 AM

Project Submission Part 3: De...Project Submission Form - P...Website Traffic DataUntitled2.ipynb - Colaborato...ChatGPT

colab.research.google.com/drive/1GPRR_cMHdvMIVX7LnzNjDt1Vxoy-JbGX#scrollTo=boGSQ3izBY82

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

import matplotlib.pyplot as

Original Dataset:

	Date	Page	Visits
0	2023-08-07	about	39
1	2023-05-31	products	38
2	2023-10-16	products	65
3	2023-12-22	about	60
4	2023-10-27	blog	71

Data for the 'home' page:

	Date	Page	Visits
6	2023-05-21	home	18
11	2023-09-29	home	68
27	2023-02-03	home	75
28	2023-10-11	home	22
30	2023-12-08	home	46

Daily Total Visits:

	Date	Visits
0	2023-01-01	215
1	2023-01-02	143
2	2023-01-03	42
3	2023-01-05	184
4	2023-01-06	91

Most Visited Page: products

Sorted Dataset by Date:

	Date	Page	Visits
704	2023-01-01	blog	82

1s completed at 7:29 AM

Project Submission Part 3: De xProject Submission Form - Pl xWebsite Traffic Data xUntitled2.ipynb - Colaborato xChatGPT x

colab.research.google.com/drive/1GPRR_cMHdvMIVX7LnzNjDt1Vxoy-JbGX#scrollTo=ENMiwW5qCj-8

Untitled2.ipynb

CommentShare

FileEditViewInsertRuntimeToolsHelpAll changes saved

+ Code+ Text

✓ [6] 0s

25% 25.000000
50% 47.000000
75% 72.000000
max 99.000000

Missing Values:
Date 0
Page 0
Visits 0
dtype: int64

1s

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset from the CSV file
df = pd.read_csv("website_traffic_dataset.csv")

# Group data by page and calculate the total visits for each page
page_visits = df.groupby("Page")["Visits"].sum().reset_index()

# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(page_visits["Page"], page_visits["Visits"])
plt.xlabel("Page")
plt.ylabel("Visits")
plt.title("Website Traffic by Page")
plt.xticks(rotation=45) # Rotate x-axis labels for better readability
plt.show()
```

✓ 1s completed at 7:29 AM



Inbox - monishavenugopal5 xRegistration successful - ven xMy IBM xPro Planet Person App xUntitled3.ipynb - Colaborat x

colab.research.google.com/drive/1HuBnwQdUGVhHsNWGBENQXmB2I9LW_jVt

Colab icons

CommentShareSettingsProfile

FileEditViewInsertRuntimeToolsHelpAll changes saved

+ Code + Text

RAMDisk

3s

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix

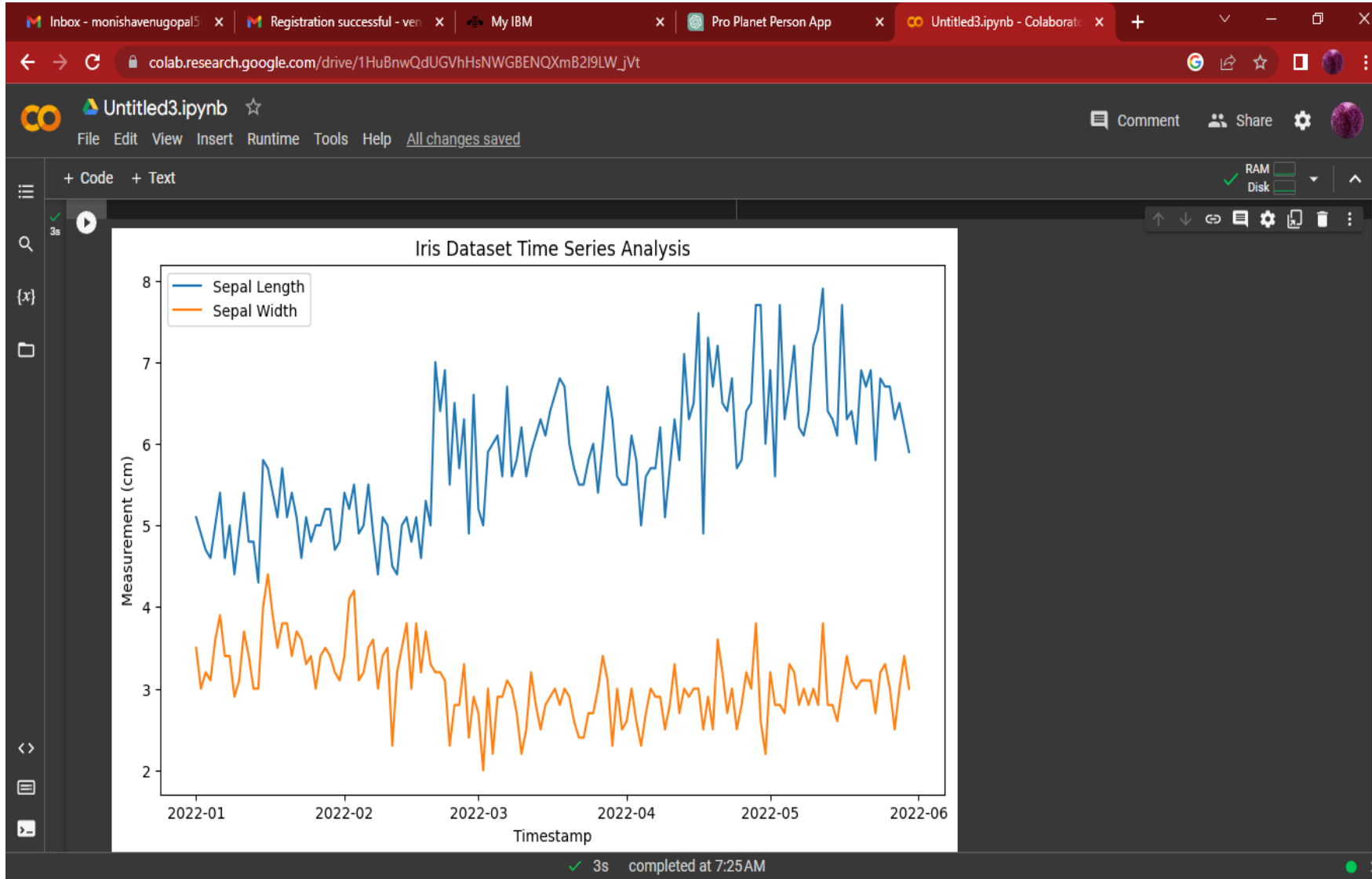
# Load the Iris dataset
from sklearn.datasets import load_iris
iris = load_iris()
data = pd.DataFrame(data=iris.data, columns=iris.feature_names)
target = pd.Series(iris.target)

# Time Series Analysis (assuming the data has a timestamp column)
data['timestamp'] = pd.date_range(start='2022-01-01', periods=len(data), freq='D')

# Plotting a time series
plt.figure(figsize=(10, 6))
plt.plot(data['timestamp'], data['sepal length (cm)'], label='Sepal Length')
plt.plot(data['timestamp'], data['sepal width (cm)'], label='Sepal Width')
plt.xlabel('Timestamp')
plt.ylabel('Measurement (cm)')
plt.title('Iris Dataset Time Series Analysis')
plt.legend()
plt.show()

# User Segmentation (clustering using K-Means)
```

3s completed at 7:25 AM



conclusion

This problem statement outlines the common challenges faced by businesses and website owners and sets the stage for a website traffic analysis project to provide solutions and recommendations.

A large, horizontal, light pink brushstroke shape with irregular, feathered edges, centered on a white background. The shape is wider on the right side and tapers slightly towards the left.

Thank you