**Website traffic analysis**

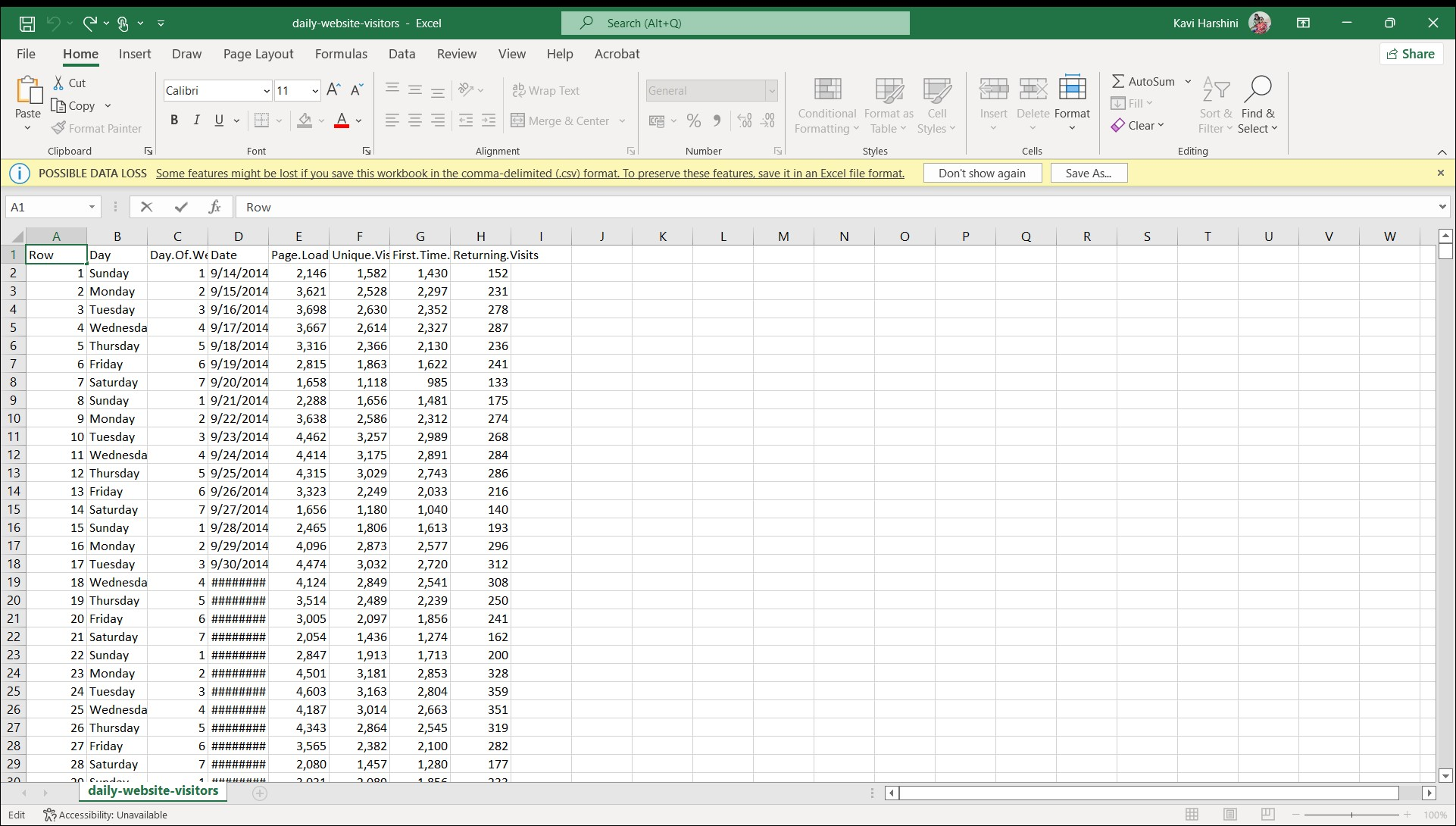
**Project: phase1**

**Project Title:** website traffic analysis

**Project Definition:**

* **The project involves analyzing website traffic data to gain insights into user behavior, popular pages, and traffic sources. The goal is to help website owners enhance the user experience by understanding how visitors interact with the site. This project encompasses defining the analysis objectives, collecting website traffic data, using IBM Cognos for data visualization, and integrating Python code for advanced analysis.**

**Dataset Link**: <https://www.kaggle.com/datasets/bobnau/daily-website-visitors>



**1.Collect Data:**

**•If you have access to server logs, you can parse them to extract information about website visits.**

**•If you use Google Analytics, you can use the Google Analytics API to retrieve data programmatically.**

**2.Data Processing:**

**•Parse and clean the data to extract relevant information like page views, user sessions, referral sources, etc.**

**•Calculate metrics like page views, unique visitors, bounce rate, and average session duration.**

**3.Visualization (Optional):**

**•Use libraries like Matplotlib or Seaborn to create charts and graphs to visualize the data.**

**4.Reporting (Optional):**

**•Generate reports or dashboards to present the analyzed data. Tools like Jupyter Notebook or web frameworks like Flask or Django can be helpful here.**

**1.Collect Data (Parsing Server Logs):**

**import re**

**log\_file\_path = "access.log"**

**def parse\_log(log\_file):**

**with open(log\_file, "r") as file:**

**log\_data = file.readlines()**

**log\_entries = []**

**for line in log\_data:**

**match = re.match(r'(\S+) (\S+) (\S+) \[([\w:/]+\s[+\-]\d{4})\] "(\S+) (\S+) (\S+)" (\d{3}) (\d+)', line)**

**if match:**

**entry = {**

**"ip": match.group(1),**

**"user": match.group(2),**

**"timestamp": match.group(4),**

**"method": match.group(5),**

**"url": match.group(6),**

**"status\_code": match.group(8),**

**"bytes\_sent": match.group(9)**

**}**

**log\_entries.append(entry)**

**return log\_entries**

**log\_entries = parse\_log(log\_file\_path)**

**2.Data Processing:**

**from collections import Counter**

**def analyze\_traffic(log\_entries):**

**page\_views = len(log\_entries)**

**unique\_ips = len(set(entry["ip"] for entry in log\_entries))**

**return page\_views, unique\_ips**

**page\_views, unique\_ips = analyze\_traffic(log\_entries)**

**print(f"Page Views: {page\_views}")**

**print(f"Unique Visitors: {unique\_ips}")**

**This is just a basic example, and real-world website traffic analysis can be much more complex depending on your requirements and the data sources you have available.**

**For more sophisticated analysis, you may want to consider using specialized web analytics tools or platforms like Google Analytics, Matomo, or Adobe Analytics, which provide APIs and documentation for data retrieval and analysis.**