

WEB TRAFFIC ANALYTICS

A report submitted for the course of
Application Development_Data Analytics Explore
III B. Tech II Semester

by

S. Hema Swaraj - 2111CS030041
G. Kranthi Varma - 2111CS030048
M. Raghavendra - 2111CS030083

Under the esteemed guidance of

Ms. K Lakshmi Madhuri
(Assistant Professor)



Department of Data Science

Malla Reddy University

Maisammaguda, Dulapally,
Hyderabad, Telangana 500100
www.mallareddyuniversity.ac.in

2024-25

CERTIFICATE

This is to certify that this bonafide record of the Application Development entitled **Web Traffic Analytics** submitted by **Mr. S. Hema Swaraj (2111CS030041)**, **Mr. G. Kranthi Varma (2111CS030048)**, **Mr. M. Raghavendra (2111CS030083)** of III year II Sem to the Malla Reddy University, Hyderabad. This bonafide record of work carried out by us under the guidance of our supervision. The contents of this report, in full or in parts, have not been submitted to any other Organization for the award of any Degree.

Internal Guide:
K. Lashmi Madhuri
Assistant Professor

Head of the Department
Dr.GS Naveen Kumar
CSE (Data Science)

External Examiner

Date:

ABSTRACT

Website traffic refers to web users who visit a website. Web analytics involves collecting, processing, visualising web data to enable critical thinking about how users interact with a web application. Web traffic is measured in visits, sometimes called "sessions," and is a common way to measure an online business effectiveness at attracting an audience.

Web analytics is the technology and method for the collection, measurement, analysis and reporting of websites and web applications usage data. Web analytics has been growing ever since the development of the World Wide Web. It has grown from a simple function of HTTP (Hypertext Transfer Protocol) traffic logging to a more comprehensive suite of usage data tracking, analysis, and reporting.

User clients, especially web browsers, generate significant data while users read and interact with webpages. The data provides insight into how visitors use the site and why they stay or leave. The key concept to analytics is learning about your users so you can improve your web application to better suit their needs.

CONTENTS

Certificate		I
Abstract		II
Contents		III
List of Figures		IV
List of Tables		IV
Chapter 1	Introduction	1
	1.1 Introduction	1
	1.2 Software Requirements	...
Chapter 2	Literature Survey	...
Chapter 3	Methodology	...
Chapter 4	Results	...
Chapter 5	Conclusions and Future Scope of Study	...
References		...

LIST OF FIGURES

Figure	Title	Page
1.1	Python Random	2
1.2	Python CSV	3
1.3	Power BI	3
3.1	Dataset	6
3.2	Power-Bi Architecture	7
3.3	Area Chart	9
3.4	Bar and Column Charts	9
3.5	Multi Row	9
3.6	Single Number	10
3.7	Combo Charts	10
3.8	Decomposition Tree	10
3.9	Doughnut Charts	11
3.10	Line Charts	11
3.11	Pie Charts	11
3.12	Slicers	12
4.1	Program Execution	13
4.2	Visualisation (1)	13
4.3	Visualisation (2)	14

LIST OF TABLES

Table	Title	Page
2.1	Terms and Definitions	5

LIST OF ABBREVIATIONS

ANNs	Artificial Neural Networks
BM	Bayesian Models
DM	Deep Learning
DR	Dimensionality Reduction
DT	Decision Trees
EL	Ensembles Learning
IBM	Instance Based Models
SVMs	Support vector Machines

CHAPTER 1

1.1: INTRODUCTION

Web traffic is the data sent and received by visitors to a website. Since the mid-1990s, web traffic has been the largest portion of Internet traffic. Sites monitor the incoming and outgoing traffic to see which parts or pages of their site are popular and if there are any apparent trends, such as one specific page being viewed mostly by people in a particular country.

Web traffic Analytics refers to collecting data about who comes to your website and what they do when they get there. That data is crucial to building effective sales and marketing strategies.

Website analytics provide insights and data that can be used to create a better user experience for website visitors. Understanding customer behavior is also key to optimizing a website for key conversion metrics.

Web analytics is the technology and method for the collection, measurement, analysis and reporting of websites and web applications usage data (Burby & Brown, 2007). Web analytics has been growing ever since the development of the World Wide Web. It has grown from a simple function of HTTP (Hypertext Transfer Protocol) traffic logging to a more comprehensive suite of usage data tracking, analysis, and reporting. The web analytics industry and market are also booming with a plethora of tools, platforms, jobs, and businesses. The market was projected to reach 1 billion in 2014 with an annual growth rate more than 15% (Lovett, 2009).

The popularity of the World-Wide Web (also called WWW, or the Web) has increased dramatically in the past few years. Data on the web is rapidly increasing day by day. Web is an open medium Today, WWW traffic is one of the dominating components of Internet traffic. There are many reasons behind this explosive growth in Web traffic. These reasons include: the ease of use of the Web, the availability of graphical user interfaces for navigating the Web, the availability of editors and support tools for creating and —publishing Web documents, the machine-independent nature of the languages and protocols used for constructing and exchanging Web documents and a continuing exponential increase in the number of Internet hosts and users. The Web Server data is actually the user logs that are generated on the Web

Server. These logs enable the analyst to keep a track of and the user's who visit the website. The process of server performance begins with the collection of log files spanning some analysis time period. To understand traffic trends there is need to collect logs over a certain period. The phenomenal growth in Web traffic has sparked much research activity on improving the World-Wide Web. Much of this recent research activity has been aimed at improving Web performance. The key performance factors to consider are how to reduce the volume of network traffic produced by Web clients and servers, and how to improve the response time for WWW users.

A. Web Traffic

The web traffic starts with the high-level activities such as clicking a link and increases with low-level activities such as travelling through network switches and cables. In other words, Web traffic is usually initiated by users via the use of web browsers. It begins with a click to access a URL. Traffic flow starts with a mouse click, which sends browser information to a server that uses predetermined rules and methods to obtain user browser requests. Based on these rules, the server then decides what action is needed. Now a days, the web traffic is increases enormously because there is continuous increase of Internet users each year is motivating online shop, gambling site, and botnet owners to take control of users moves to point them to their sites. Therefore, there is need of web traffic analysis tools. These tools handle and categorize the traffic and increase the workload handling capacity of the web server.

B. LOGS

Web server logs stores click stream data which can be useful for web traffic analysis . They are plain text (ASCII) files which contain information about User Name, IP Address, Time Stamp, Access Request, URL that Referred, error codes (if any) etc. and generally reside in the web servers. Traditionally there are four types of server logs: Transfer Log, Agent Log, Error Logo. The Transfer and the Agent Log are said to be standard whereas the error and referrer log are considered optional as they may not be turned on. Every log entry records the traversal from one page to another, storing user IP number and all the related information. If logs are utilized properly, it can be very useful in turning the websites visitors into customers especially in case of an e-commerce website. It guides the analyst in determining the navigational pattern of the user i.e. which pages are frequently visited by the user, the kind of errors that user gets, etc. A variety of tools are available that take the web access logs as an input and generate the reports as an output. These tools provide us with all sorts of information starting from how

many hits the site getting to the number of visitors accessing the site, the browsers that they use the length of their stay, and much more. Some of the tools that are available are:

1) Google Analytics: It is a free utility provided by Google which helps in keeping a track of unique visitors. It also helps in determining which marketing packages are offering the best . For using this tool, installation is not required, only requires a Google account. Email report facility is available in Google analytics.

2) AWStats : It is available free of cost. This tool works as a CGI Script or from command line. It displays all sorts of information that the log contains.

3) Web Log Expert: Yet another log Azer tool that provides thorough analysis of the web access logs. It provides the users with specific and precise information about user statistics. It supports log files extracted from Apache and IIS. For using this tool, there is no need of creating any account but profile creation is required.

4) Analog: This is an easy to use and install freely available log analysis tool. It is extremely fast, highly scalable, works on any operating system and an easy to install tool.

WEB SERVER LOG ANALYZER

Automating the analysis of server logs is essential to allow using the logs as a proactive administrator tool. Log analyzer is 3- tier architecture based software that parses through the log files generated by whichever web server follows the standard web server logging standards. Analyse parsed data and categorize them into meaningful reports that can be read by the user for administration or monitoring purpose. A software application designed to parse a log file, which typically contains raw collected data, and convert it to an easy-to-read and understand form. These log files can be very large and are very detailed on which files were requested from our web server. Every time a page, or image, or movie, or any other kind of file is downloaded off of your web server, the date/time and IP address of the requestor is logged in the web server's log file. Web logs can provide information about which user groups access the website, which resources are viewed the most, and the links users follow to arrive at the site. The web server log Analyzer produces reports with summary statistics about website.

1.2: SOFTWARE REQUIREMENT

We have used Python and Microsoft Power BI to generate sample data and represent the data respectively. The list of modules from python that has been used are:

1. Random Module
2. CSV Module

PYTHON:

Python is an interpreted, high-level, general purpose programming language created by Guido Van Rossum and first released in 1991, Python's design philosophy emphasises code Readability with its notable use of significant White space. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage collected. It supports multiple programming paradigms, including procedural, object oriented, and functional programming.

1.Random Module:

The Python Random module is a built-in module for generating random integers in Python. These numbers occur randomly and does not follow any rules or instructions. We can therefore use this module to generate random numbers, display a random item for a list or string, and so on.



fig 1.1: Python Random

2. CSV Module :

The CSV module implements classes to read and write tabular data in CSV format. It allows programmers to say, “write this data in the format preferred by Excel,” or “read data from this file which was generated by Excel,” without knowing the precise details of the CSV format used by Excel



fig 1.2: Python CSV

POWER BI:

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data might be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses.

Power BI consists of several elements that all work together, starting with these three basics:

- A Windows desktop application called Power BI Desktop.
- An online software as a service (SaaS) service called the Power BI service.
- Power BI Mobile apps for Windows, iOS, and Android devices.

These three elements—Power BI Desktop, the service, and the mobile apps—are designed to let you create, share, and consume business insights in the way that serves you and your role most effectively.

Beyond those three, Power BI also features two other elements:

- Power BI Report Builder, for creating paginated reports to share in the Power BI service. Read more about paginated reports later in this article.
- Power BI Report Server, an on-premises report server where you can publish your Power BI reports, after creating them in Power BI Desktop.



fig 1.3: Power BI

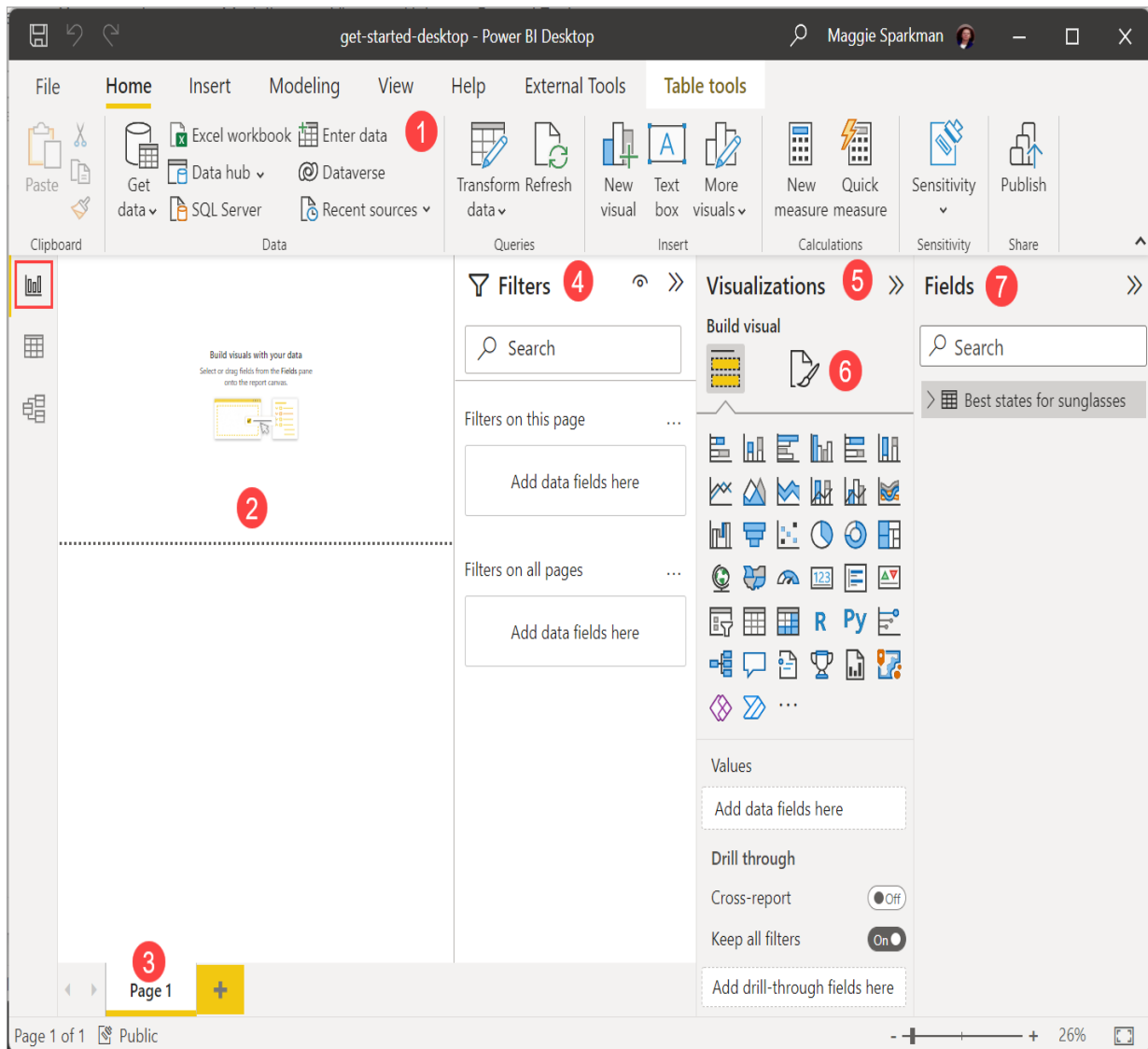


fig 1.4: Power BI view

CHAPTER 2

2: LITERATURE SURVEY

Log files have been used to keep track of web requests since World Wide Web emerged and the first widely used browser Mosaic was released in 1993. One of the pioneers of web log analysis was Web Trends, a Portland, Oregon based company, which conducted website analytics using data collected from web server logs. In the same year, Web Trends created the first commercial website analytics software. In 1995, Dr. Stephen Turner created Analog, the first free log file analysis software. In 1996, Web Side Story offered a hit counter as a service for websites that would display a banner. Web server logs have some limits in types of data collected. For example, they could not provide information about visitors' screen sizes, user interactions with page elements, mouse events such as clicking and hovering, etc. The new

technique of page tagging is able to overcome the limitation and has become more popular recently.

The fundamental basis of web analytics is collection and analysis of website usage data. Today, web analytics is used in many industries for different purposes, including traffic monitoring, e-commerce optimization, marketing/advertising, web development, information architecture, website performance improvement, web-based campaigns/programs, etc

The Web Analytics Association (now the Digital Analytics Association) was founded in 2004. This formalised the creation of more stabled and essential tools. They developed a comprehensive and in-depth body of knowledge to help websites optimise and new methods and tools were created to report massive amounts of data. The rise of search engine marketing (SEM) and social media platforms meant that businesses needed to keep up with multiple traffic channels, understand trends, and make sure that they could optimise for different channels. The number of individuals connected to the internet grew globally and marketers kept finding ways to take advantage.

Web analytics are measured by software that tracks website visitors' mouse clicks and information requests. The data are stored by Google and can be compared over time to help Web managers improve the effectiveness of websites, and managers make decisions about campaign effectiveness. The data gathered by Google Analytics can be used to determine which pages on an organisation's website are the most popular or most accessed, what type of information visitors to the site are interested in.

2.1: KEYWORDS USED IN WEB TRAFFIC ANALYTICS

S.No	KEYWORDS	DEFINITION
1.	Page View	The number of times a page (an analyst-definable unit of content) was viewed.
2.	Visits	A visit is an interaction, by an individual, with a website consisting of one or more requests for an analyst-definable unit of content (i.e. “page view”).
3.	Unique Visits	Each individual is counted only once in the unique visitor measure for the reporting period. The number of visits by different numbers of persons.
4.	Paid Traffic	Traffic for which you need to pay, e.g., Google AdWords.
5.	Organic Traffic	The traffic coming through all search engines (Google, Yahoo, Bing....)
6.	Direct Traffic	The traffic coming directly to your website.
7.	Social Traffic	The traffic coming through all social media platforms
8.	Referral Traffic	The traffic coming through where your website is linked.
9.	Bounce Rate	The percentage of all sessions on your site in which users viewed only a single page
10.	Conversion Rate	The percentage of users who take a desired action.

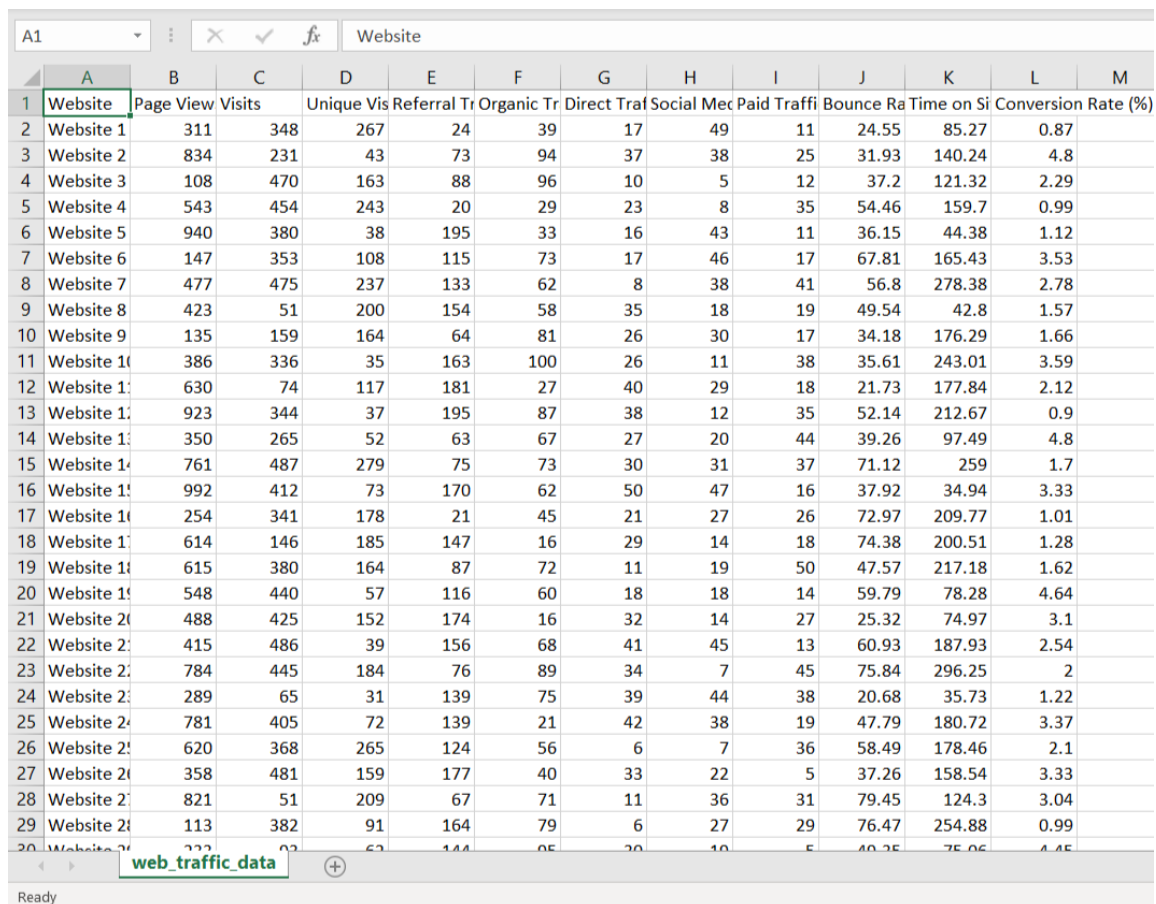
Table: 2.1 *Terms and Definitions*

CHAPTER 3

3: METHODOLOGY :

3.1: GETTING A DATASET

Since there are no sample web traffic datasets available for experimentation and representation. We have used Python libraries to generate the data containing different types of traffic for fifty different websites.



	A	B	C	D	E	F	G	H	I	J	K	L	M
	Website	Page View	Visits	Unique Vis	Referral Tr	Organic Tr	Direct Traf	Social Mec	Paid Traffi	Bounce Ra	Time on Si	Conversion Rate (%)	
2	Website 1	311	348	267	24	39	17	49	11	24.55	85.27	0.87	
3	Website 2	834	231	43	73	94	37	38	25	31.93	140.24	4.8	
4	Website 3	108	470	163	88	96	10	5	12	37.2	121.32	2.29	
5	Website 4	543	454	243	20	29	23	8	35	54.46	159.7	0.99	
6	Website 5	940	380	38	195	33	16	43	11	36.15	44.38	1.12	
7	Website 6	147	353	108	115	73	17	46	17	67.81	165.43	3.53	
8	Website 7	477	475	237	133	62	8	38	41	56.8	278.38	2.78	
9	Website 8	423	51	200	154	58	35	18	19	49.54	42.8	1.57	
10	Website 9	135	159	164	64	81	26	30	17	34.18	176.29	1.66	
11	Website 10	386	336	35	163	100	26	11	38	35.61	243.01	3.59	
12	Website 11	630	74	117	181	27	40	29	18	21.73	177.84	2.12	
13	Website 12	923	344	37	195	87	38	12	35	52.14	212.67	0.9	
14	Website 13	350	265	52	63	67	27	20	44	39.26	97.49	4.8	
15	Website 14	761	487	279	75	73	30	31	37	71.12	259	1.7	
16	Website 15	992	412	73	170	62	50	47	16	37.92	34.94	3.33	
17	Website 16	254	341	178	21	45	21	27	26	72.97	209.77	1.01	
18	Website 17	614	146	185	147	16	29	14	18	74.38	200.51	1.28	
19	Website 18	615	380	164	87	72	11	19	50	47.57	217.18	1.62	
20	Website 19	548	440	57	116	60	18	18	14	59.79	78.28	4.64	
21	Website 20	488	425	152	174	16	32	14	27	25.32	74.97	3.1	
22	Website 21	415	486	39	156	68	41	45	13	60.93	187.93	2.54	
23	Website 22	784	445	184	76	89	34	7	45	75.84	296.25	2	
24	Website 23	289	65	31	139	75	39	44	38	20.68	35.73	1.22	
25	Website 24	781	405	72	139	21	42	38	19	47.79	180.72	3.37	
26	Website 25	620	368	265	124	56	6	7	36	58.49	178.46	2.1	
27	Website 26	358	481	159	177	40	33	22	5	37.26	158.54	3.33	
28	Website 27	821	51	209	67	71	11	36	31	79.45	124.3	3.04	
29	Website 28	113	382	91	164	79	6	27	29	76.47	254.88	0.99	
30	Website 29	322	62	62	144	95	20	10	5	40.35	75.06	4.45	

fig 3.1: Dataset

3.2: POWER BI ARCHITECTURE:

The major building blocks of Power BI are: dashboards, reports, workbooks, datasets, dataflows, and apps. They're all organised into workspaces, and they're created on capacities.

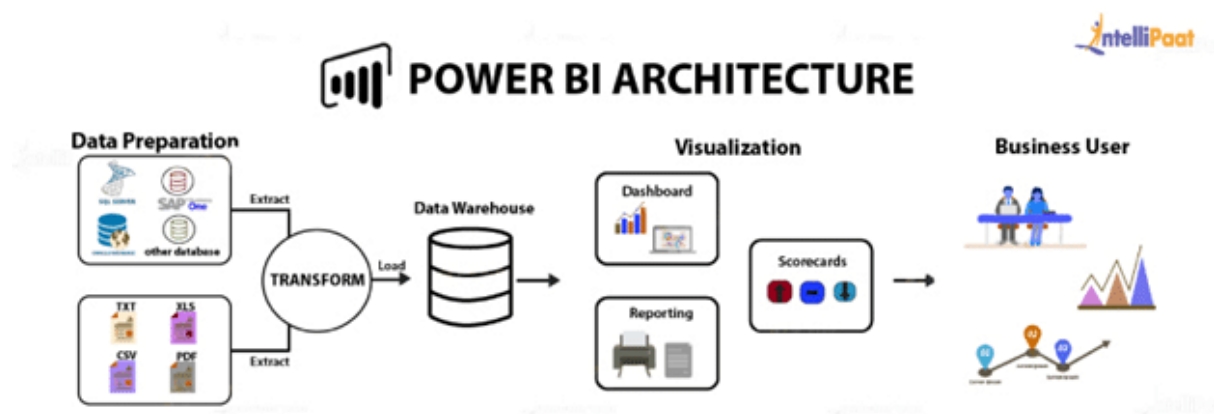


fig 3.2: Power-Bi Architecture

MS Power BI architecture consists of four major steps that explain the whole process from data sourcing to the creation of reports and dashboards. Various technologies and processes work together to get the required results with extreme precision. Let's see those steps further.

3.2.1: SOURCING DATA:

Power BI extracts data from various servers, Excel sheets, CSV files, and databases. The extracted information can be directly imported to Power BI, or a live service link is

established to receive it. If you directly import the data in Power BI, it will only be compressed up to 1 GB. Post that, you can only run



fig 3.3: Power-Bi sourcing data

3.2.2: TRANSFORMING THE DATA:

Before visualising the data, cleaning and preprocessing it should be done. This means removing useless or missing values from rows or columns. Following that, certain rules will be applied to transform and load the datasets into the warehouse.

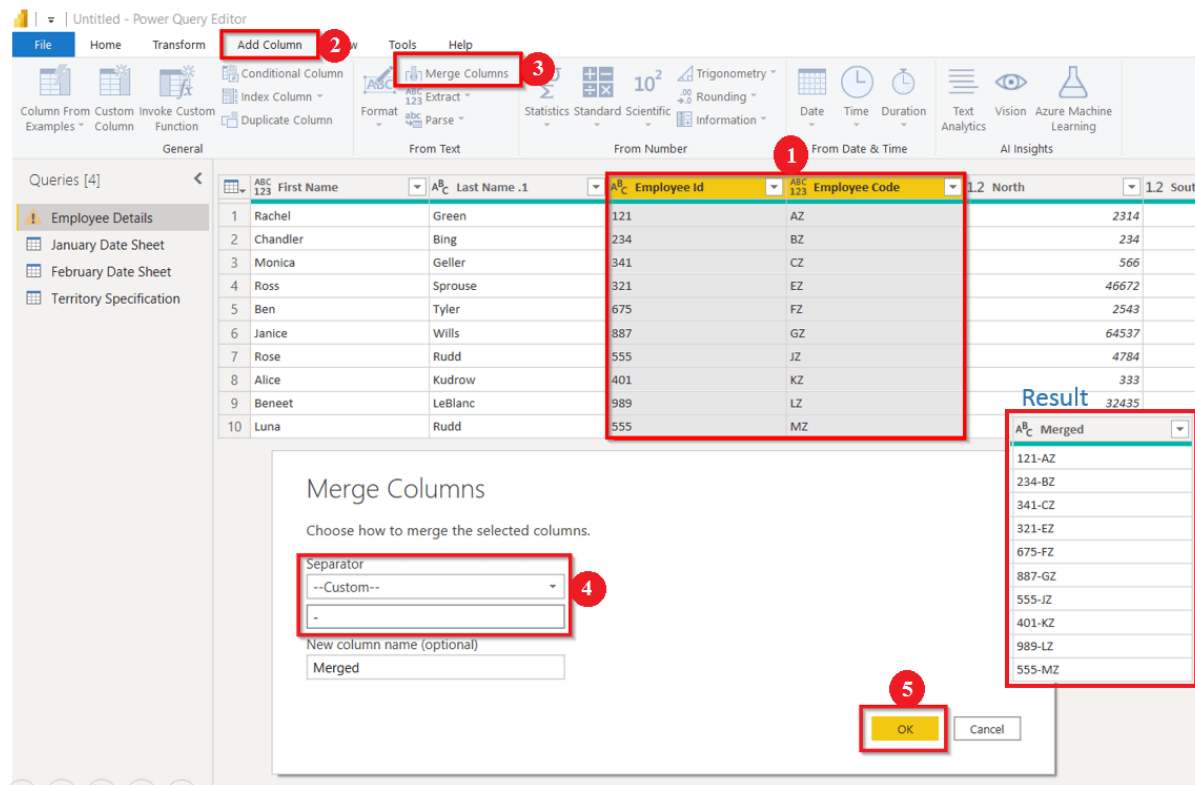


fig 3.4: Power-Bi Transforming the data.

3.2.3: REPORT AND PUBLISH:

After cleaning and transforming the data, reports will be created based on requirements. A report is a visualisation of the data with different filters and constraints presented in the form of graphs, pie charts, and other figures.

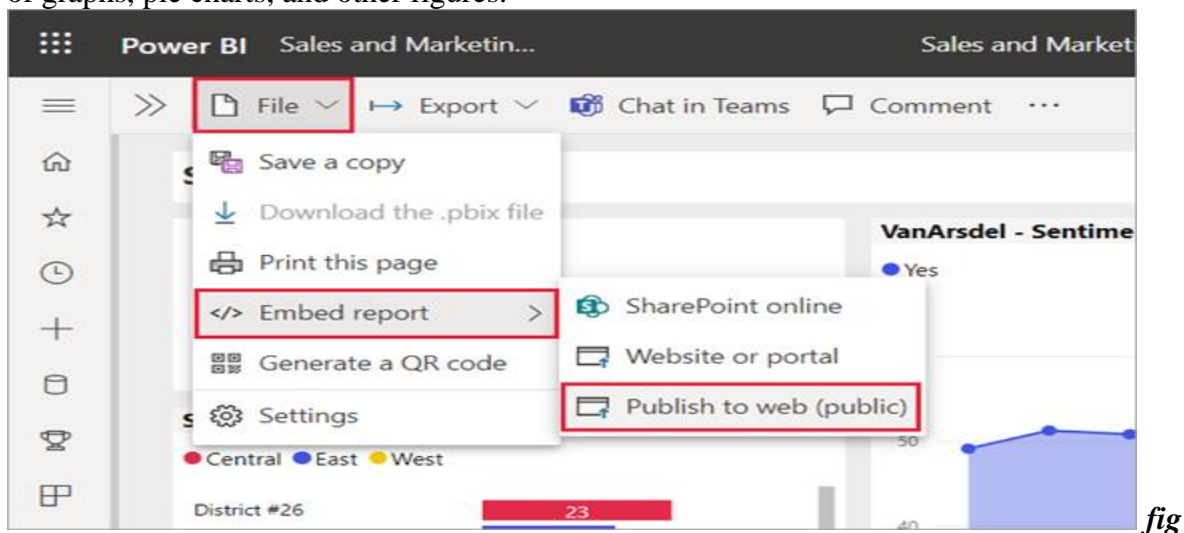


fig 3.5: Power-Bi Report and Publish

3.2.4: CREATING DASHBOARDS:

Power BI Dashboards are created by pinning individual elements or pages of live reports. Dashboards should be created after you have published your reports to the BI service. When the reports get saved, the visual maintains the filter settings chosen so that the user can apply filters and slicers.



fig 3.6: Creating Dashboards.

3.3: POWER BI VISUALISATION:

In Power BI, you can create visualisation in two ways. First is by adding from the right side pane to Report Canvas. By default, it is the table type visualisation, which is selected in Power BI. Another way is to drag the fields from the right sidebar to the axis and value axis under visualisation. You can add multiple fields to each axis as per the requirement.

3.3.1: AREA CHARTS:



fig 3.6: Area Charts

3.3.2: BAR AND COLUMN CHARTS:

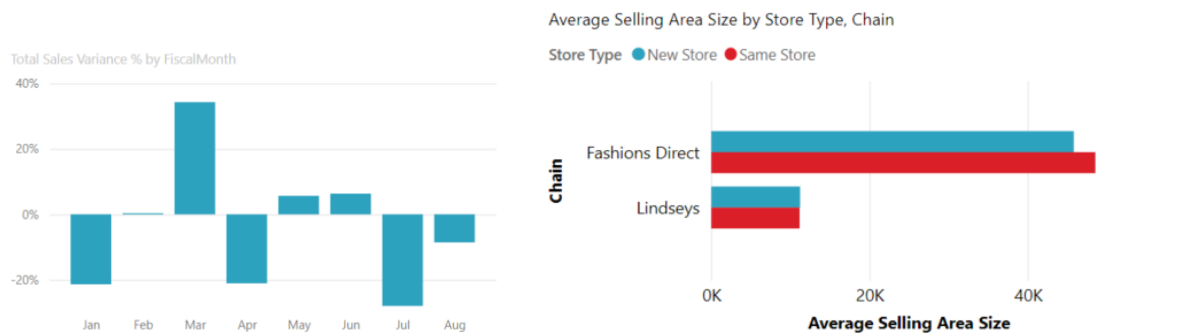


fig 3.7: Bar and Column Charts

3.3.3: CARDS:

1. MULTI ROW

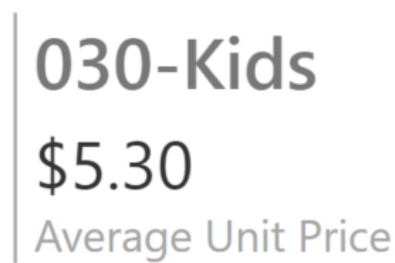


fig 3.8: Multi Row

2. SINGLE NUMBER

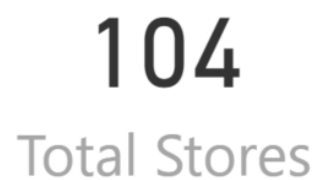


fig: 3.9: Single Number

3.3.4: COMBO CHARTS:



fig: 3.10: Combo Charts

3.3.5: DECOMPOSITION TREE:

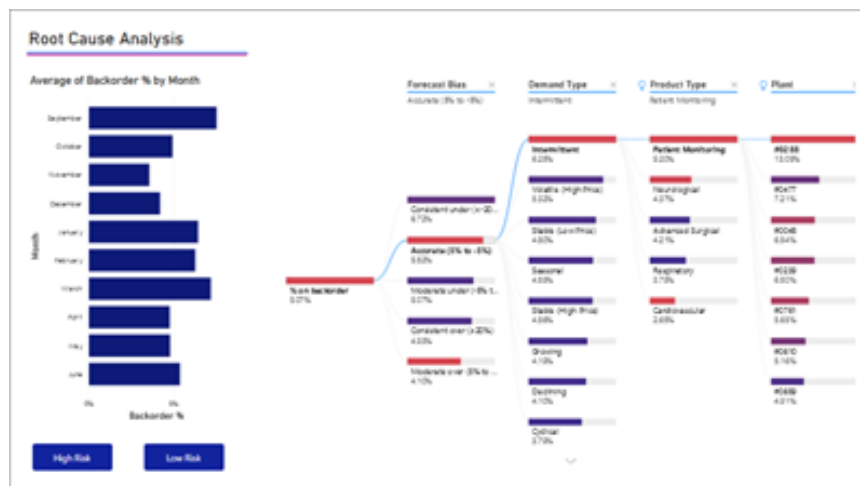


fig: 3.11: Decomposition Tree

3.3.6: DOUGHNUT CHARTS:

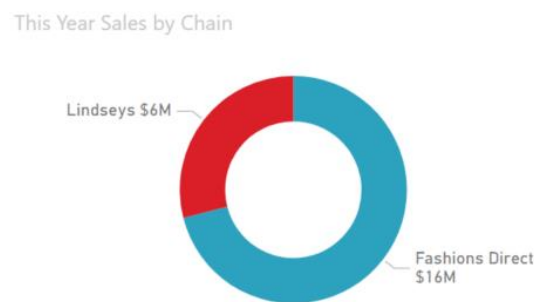


fig: 3.12: Doughnut Charts

3.3.7: LINE CHARTS:

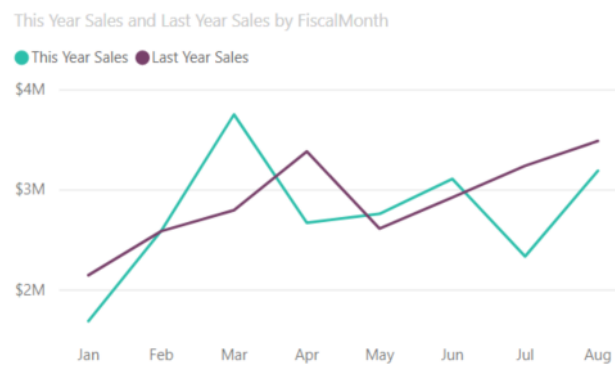


fig: 3.13: Line Charts

3.3.8: PIE CHARTS:

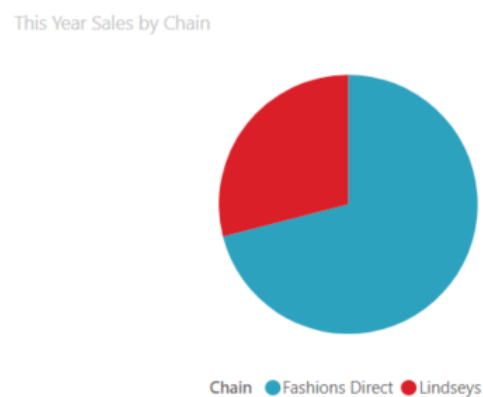


fig: 3.14: Pie Charts

3.3.9: SLICERS:

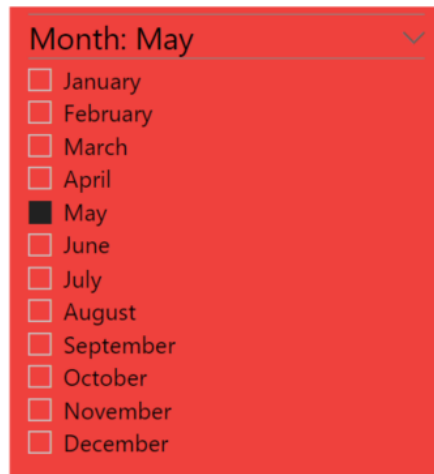


fig: 3.15: Slicers

CHAPTER 4

4. RESULTS:

We have been able to execute the python program for data generation successfully. The python program contains *RANDOM* and *CSV* modules. The generated data in the CSV format contained 50 rows and 12 columns.

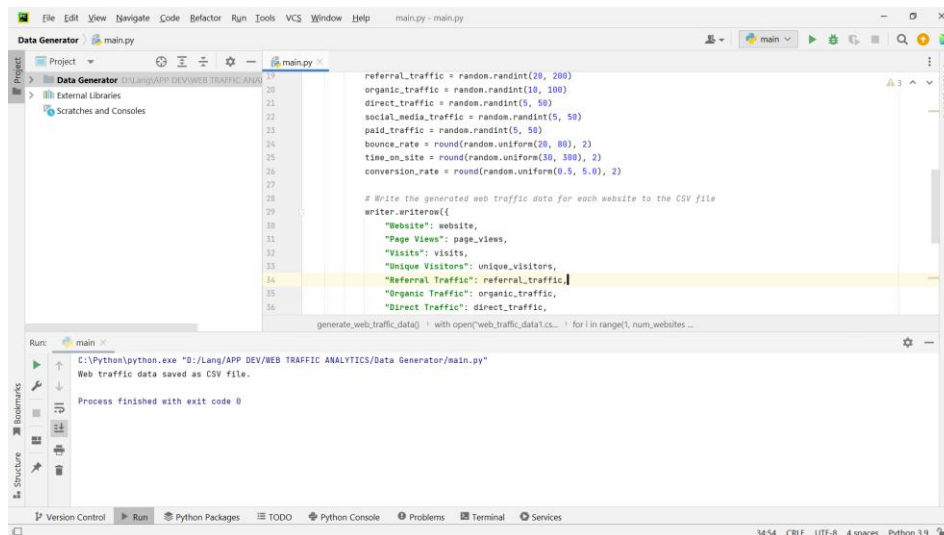


fig 4.1: Program Execution

The Data has been successfully imported and visualised in the Power-BI.

The 1st visualisation shows the Area Graph between the ‘Unique Visitors’, ‘Conversion Rate’, and the ‘Bounce Rate’.

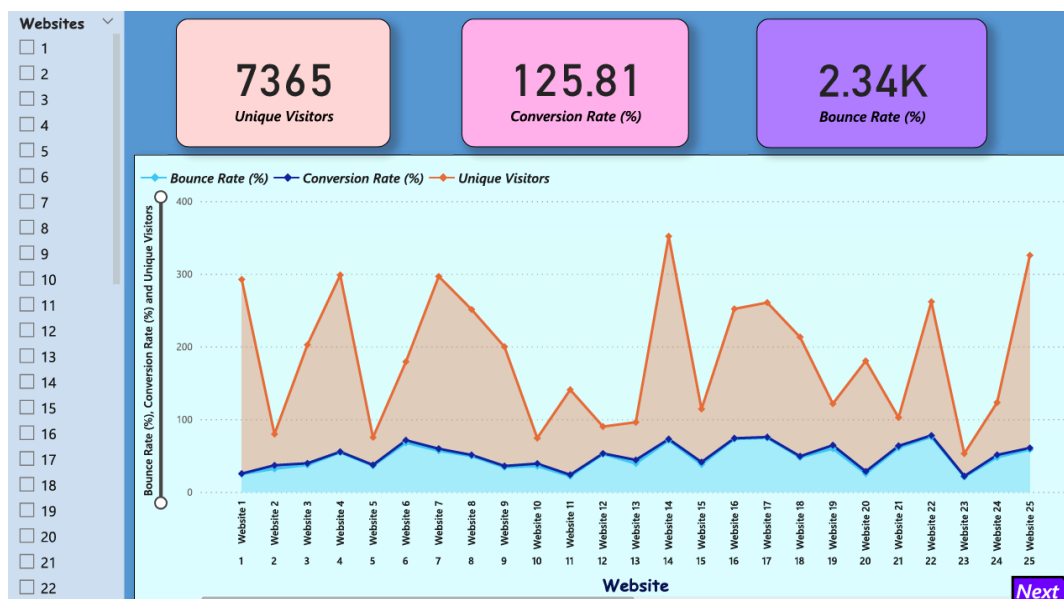


fig 4.2: Visualisation (1)

The 2nd visualisation shows Doughnut Charts of Direct Traffic and Organic Traffic & Paid Traffic, Referral Traffic and Social Media Traffic .

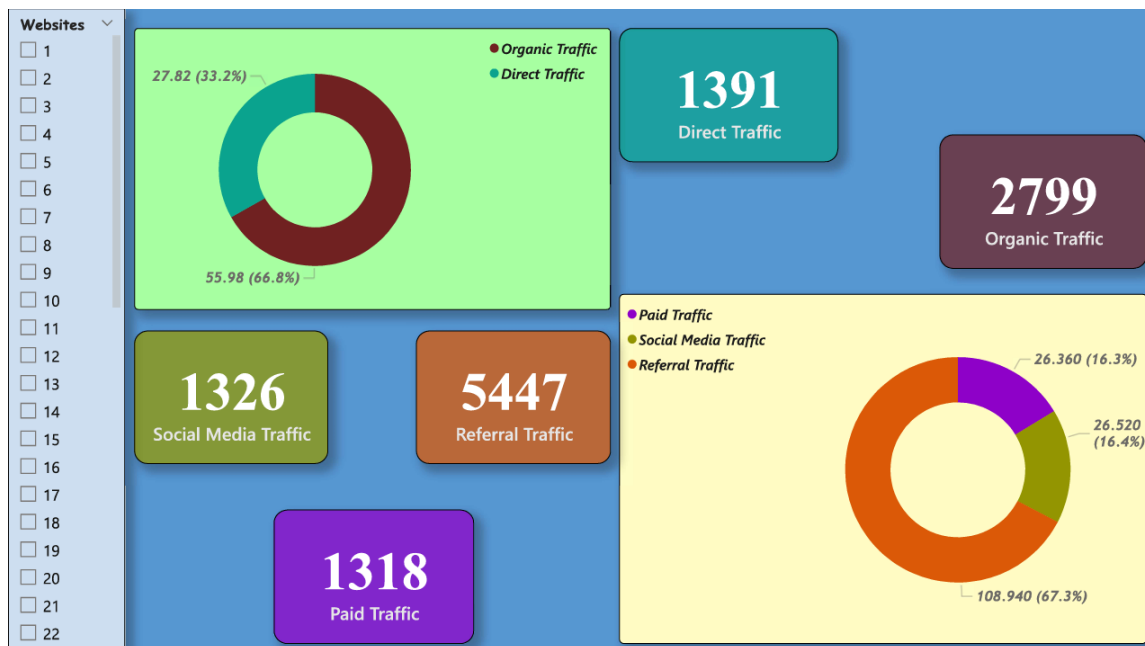


fig 4.3: Visualisation (2)

CHAPTER 5

5.CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION:

Web analytics is a field of web traffic data collection and analysis. It had gained wide adoption and become one of the important tools to help web application management and business analysis. With the recent Web 2.0 and cloud service advancements, it has quickly evolved from simple system level data logging to more comprehensive information collection and analysis. With the continuing expansion of data sources, Web/digital analytics will play an even more important role in the future.

5.2 FUTURE SCOPE:

Web/digital analytics is a rapidly growing field, and its importance is only increasing as more businesses move their operations online. As a career option, it offers a wide range of opportunities for those who are interested in data analysis, marketing, and technology. The scope of web/digital analytics is vast, and it includes everything from analysing website traffic and user behaviour to tracking social media engagement and email campaigns. With the increasing use of AI and machine learning, the field is becoming more sophisticated, and there is a growing need for professionals who can interpret and make sense of the vast amounts of data generated by online interactions. In terms of the future, web/digital analytics is expected to continue to grow and evolve. As more businesses move online, there will be an increasing demand for professionals who can help them make data-driven decisions and optimise their online presence. Additionally, as technology continues to advance, there will be new opportunities to analyse and leverage data from emerging platforms such as virtual and augmented reality. Overall, web/digital analytics is a promising career option for those who are interested in technology and data analysis.

REFERENCES

- [1] Qiu, F., Liu, Z. and Cho, J., 2005, June. Analysis of user web traffic with a focus on search activities. In *WebDB* (pp. 103-108).
- [2] Ihm, S. and Pai, V.S., 2011, November. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (pp. 295-312).
- [3] Burby, J., Brown, A. and WAA Standards Committee, 2007. *Web analytics definitions*. Washington DC: Web Analytics Association.
- [4] Zheng, G. and Peltsverger, S., 2015. *Web analytics overview*. In *Encyclopedia of Information Science and Technology, Third Edition* (pp. 7674-7683). IGI Global.
- [5] <https://activeinternetmarketing.co.uk/what-are-the-different-types-of-website-traffic/>
- [6] https://youtu.be/PnVZ7_OA7Qo
- [7] Chen, J. and Cheng, W., 2016, September. Analysis of web traffic based on HTTP protocol. In *2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 1-5). IEEE.