Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

A report on Demographic analysis using Power BI

The domain of the Project:

Data Analytics in Social and Population Studies

Under the guidance of

Mrs. Siddhika Shah

By

Polakal Shaik Affan(B.Tech, 4th year)

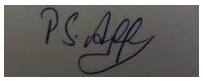
Period of the project

February 2025 to March 2025

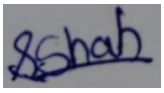
Declaration

The project titled "A report on Demographics using Power BI" has been mentored by Mrs. Siddhika Shah, organised by SURE Trust, from February 2025 to March 2025, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge the members of the team mentioned below, have worked on it successfully and enhanced their practical knowledge in the domain.

Name : Polakal Shaik Affan

Signature: 

Mentor's Name: Mrs. Siddhika Shah

Signature: 

Prof. Radhakumari
Executive Director & Founder
SURE Trust

Table of contents

Executive Summary.

Demographics analysis involves examining the characteristics of a population based on factors such as age, gender, location, income, education, and more. It helps businesses and organizations understand their audience, tailor marketing strategies, personalize customer experiences, and make informed decisions. So, if you want to learn how to perform demographics analysis, this article is for you. In this article, I'll take you through the task of Demographics Analysis with Python.

Demographics Analysis means understanding the composition of a website's audience based on various demographic factors such as age, gender, location, and interests. Expected results from demographics analysis include detailed insights into who the visitors are, how different demographic groups interact with the site, and which segments are most valuable or underrepresented.

It helps businesses tailor their content, marketing strategies, and user experience to better meet the needs and preferences of their target audience, ultimately driving engagement and conversion rates.

Introduction

The Demographic Analysis project was created to provide an easy-to-understand, data-focused view of population trends through interactive visualizations within Power BI. Based on organized demographic datasets, the project examines central factors including gender distribution, marital status, work category, income range, and education level. The overall aim was to provide these indicators through an easy-to-use dashboard that facilitates decision-making for purposes such as social planning and resource allocation. The report combines custom and native Power BI visuals like bar charts, pie charts, and a Dumbbell Bar Chart to accentuate comparative trends, such as gender disparity in employment and education, and categorical segmentation by age groups and income levels. The project scope is focused on descriptive analysis, where insights are taken from a static dataset, and is restricted to visualization and filtering capabilities without predictive modelling or live updates. One of the most important innovations is the employment of custom visuals and slicers to enable interactive engagement with the data by users, presenting the insights accordingly to meet chosen criteria such as marital status, educational level, or work class. This interactivity gives thickness to the analysis and turns the dashboard into an effective tool for stakeholders requiring rapid, visual comprehension of intricate demographic trends.

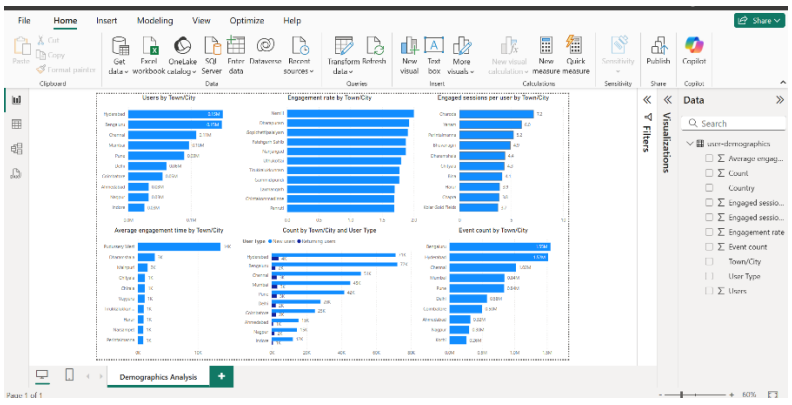
Project Objectives

The Demographic Analysis project was constructed with the central aim of comparing user engagement patterns in different towns and cities of different countries. Employing fields like Town/City, Country, Users, New Users, Engaged Sessions, Engagement Rate, Engaged Sessions per User, Average Engagement Time, and Event Count, the aim was to provide a comprehensive and interactive Power BI dashboard that graphically displays how user behaviour differs between regions. The project is to give stakeholders a view of which areas are most active in users, retain the most, and engage with most, assisting in making strategic decisions in marketing, user experience design, and local outreach. Expected results were determining high- and low-performing regions, comprehending patterns of user retention, and analysing which towns or nations exhibit the most stable user interaction. The primary outputs of this project are a dynamic Power BI dashboard that includes several types of charts—like bar graphs, pie charts, and comparative visuals—accompanied by slicers and filters that allow regional-level analysis. Not only does this project simplify raw user interaction data into easy-to-understand visual form, but it also allows stakeholders to quickly establish relevant insights to inform business or organizational strategies.

Methodology and Results

The **Demographic Analysis** project was developed using modern data analytics methodologies and business intelligence technologies to visualize and interpret user engagement across various towns and countries. The core method involved **data cleaning, transformation, and interactive visualization**, all executed within **Microsoft Power BI**, which served as both the data processing and visualization platform. Key tools and technologies used include **Power BI Desktop** for report creation, **Power Query Editor** for preprocessing the dataset, and **DAX (Data Analysis Expressions)** for calculating metrics such as engagement rate and sessions per user. The dataset, provided in CSV format, included attributes such as Town/City, Country, Users, New Users, Engaged Sessions, and Engagement Time, and was pre-collected—thus no manual or live data collection was necessary. The **project architecture** follows a simple yet effective structure: data is first imported and cleaned using Power Query, followed by the creation of data relationships (if multiple tables exist), and then dynamic dashboards are built using visual components and custom slicers to allow granular exploration by location or metric. The final output includes multiple visual pages that showcase insights such as top-performing cities, countries with the highest user retention, and average engagement times. Screenshots of the Power BI dashboard display visuals like bar charts comparisons, along with slicers that allow users to filter by country or session types for deeper insights. The project is documented and hosted on GitHub for version control and collaboration.

GitHub Repository: https://github.com/sure-trust/POLAKAL-SHAIK-AFFAN-g2-fsds/blob/main/Final%20capstone%20project/Demographics%20Analysis.pbix




```
{'High Engagement Users': 1383,
 'New Users': 4975,
 'Returning Users': 17,
 'Low User Cities': 1687,
 'Medium User Cities': 1676,
 'High User Cities': 1637}
```

The first Graph i.e., user by town/ City say Hyderabad, Bengaluru, and Chennai are the top three cities with the highest number of users, significantly outpacing other cities. There is a steep drop-off in user count after the top few cities, which indicates a concentration of users in major urban centres.

The Second graph Engagement by town/city, key findings indicate that Nemili, Dharapuram and Gopichettipalaiyam exhibit the highest engagement rates, suggesting that users in these cities are more actively interacting with the website's content.

Engaged session per user by town/city says key findings indicate that Charoda and Yanam have the highest engaged sessions per user, suggesting that users in these cities interact frequently with the website.

Average engagement time key findings include Pudussery west, Dharamshala and Mainpuri showing the highest average engagement times, indicating that users in these cities spend more time engaged on the website.

Count by town/city and user type data reveals that the majority of users in most cities are new users, with Hyderabad, Bengaluru, and Chennai having the highest counts. This suggests that these cities are experiencing significant growth in attracting first-time visitors. Conversely, the proportion of returning users is much smaller across all cities, indicating a potential area for improving user retention strategies.

Event count by town/city says Hyderabad and Bengaluru lead significantly with the highest event counts, indicating a high level of user interactions in these cities. Chennai and Mumbai also exhibit substantial event counts, demonstrating active user engagement. The cities towards the lower end of the chart, such as Bhubaneswar, Jaipur, and Istanbul, have relatively lower event counts, suggesting less user interaction.

Learning and Reflection

During the Demographic Analysis project, I was able to improve my technical and analytical abilities, particularly in data visualization and business intelligence. Using Power BI, I experienced working hands-on with cleaning data using Power Query, developing dynamic visualizations using DAX, and designing custom reports to review important measures like user engagement, session rates, and average engagement time in different towns and countries. I also improved my knowledge of how to properly design data models, establish relationships, and create interactive dashboards that give stakeholders actionable information. At a management level, I improved my project management skills by organizing and conducting the whole project independently, from data gathering to report presentation. This involved time management, problem-solving, and the capacity to adjust to challenges on my own. Overall, the project was a worthwhile learning experience that not only enhanced my technical skills in Power BI but also enhanced my skill set in managing and delivering a full-cycle data analytics project from beginning to end. The feeling of satisfaction after witnessing the final interactive dashboard materialize was truly fulfilling, and it gave me a greater appreciation for the strength of data-driven decision-making.

To effectively allocate the budget using these user segments, we should focus on strategies that maximize returns by targeting specific groups. Here's how we can approach budget allocation based on each segment:

1. **High Engagement Users:** These users are already highly engaged, meaning they have a high likelihood of converting or purchasing. We can use a substantial portion of the budget for this segment to maintain and enhance their engagement. We can offer personalized ads, loyalty programs, and exclusive offers to this segment.
2. **New Users:** These users represent potential growth. Effective onboarding can turn them into regular, engaged users. We can invest in welcoming campaigns, introductory offers, and onboarding programs for this segment. The focus should be on making their first experiences positive to drive retention.
3. **Returning Users:** While the number of returning users is small, they represent loyalty and consistency. We can offer retention strategies, such as personalized recommendations and rewards for continued use to this segment of users.
4. **Low User Cities:** These cities have untapped potential. Awareness campaigns can increase the user base. We can allocate a smaller portion of the budget to brand awareness campaigns and introductory offers to attract new users from these cities.
5. **Medium User Cities:** These cities show moderate potential and can be nurtured to become high-user cities. We can allocate more budget to this segment in the form of targeted ads to convert them into cities with high users.
6. **High User Cities:** These cities already have a large user base with high engagement potential. Similar to the high-engagement users, we can use a substantial portion of the budget for this segment as well to maintain and enhance their engagement.

So, this is how we can make data-driven decisions using techniques like demographics analysis for budget allocation.

The Demographic Analysis project was created to compare user engagement rates between various towns and nations using influential metrics like users, new users, engaged sessions, and engagement rate. The overall aim was to take raw demographic information and present it in a meaningful way that would assist stakeholders in learning about regional user activity, retention, and engagement patterns. Through the creation of an interactive Power BI dashboard, the project had successfully reached its objective of presenting a dynamic interface for discovering such insights via bar charts, pie charts, and specialized visuals such as the Dumbbell Bar Chart. One of the most important successes was creating a dashboard that enables users to filter and segment the data by town, country, and engagement metrics, so high-performing regions and areas to focus on are easily identifiable. The project also emphasized the significance of clean data and good visualization methods to explain complicated trends clearly. In the future, the scope of this project involves integrating real-time data feeds to dynamically update engagement metrics, adding predictive analytics to predict user behaviour trends, and broadening the analysis to incorporate other demographic variables such as age, gender, or income level. Additionally, incorporating more sophisticated machine learning models would enhance the accuracy of user retention predictions and allow for more insightful analysis of user engagement drivers.