Case Study Report



Data Analytics with Power BI

**“AN ANALYSIS OF UNEMPLOYMENT IN REPUBLIC OF INDIA (DATA ANALYSIS WITH POWER BI)”**

**“ MEENAKSHI CHANDRASEKARAN COLLEGE OF ARTS & SCIENCE ”**

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| 60EAD9982873695B4A346F7C57DEF51A | M.SAMEEHA |

|  |  |
| --- | --- |
|  |  |
|  | Trainer Name |
|  | Master Trainer |

**ABSTRACT**

This case study delves into the pressing issue of unemployment in the Republic of India, utilizing data analytics methodologies with a focus on Power BI. By examining a diverse array of socio-economic indicators and demographic variables, the study aims to uncover underlying trends and patterns contributing to unemployment rates across different regions and population segments.

Employing a robust dataset spanning multiple years, the analysis employs Power BI's advanced visualization and analytical capabilities to explore correlations and identify key drivers of unemployment. Through interactive dashboards and dynamic visualizations, the study provides insights into the impact of factors such as education, industry composition, and regional disparities on employment opportunities.

Key findings shed light on the complex interplay of socio-economic factors influencing unemployment rates in India, offering valuable insights for policymakers, researchers, and practitioners. By harnessing the power of data analytics with Power BI, this case study demonstrates the potential for informed decision-making and targeted interventions to address the challenges of unemployment and foster inclusive economic growth in the Republic of India.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 4 |
| 2 | Chapter 2: Services and Tools Required | 9 |
| 3 | Chapter 3: Project Architecture | 11 |
| 4 | Chapter 4: Modeling and Result | 13 |
| 5 | Conclusion | 22 |
| 6 | Future Scope | 23 |
| 7 | References | 25 |
| 8 | Links | 25 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Problem Statement**

Despite India’s economic growth and development strides, persistent challenges with unemployment continue to hinder inclusive prosperity and socio-economic advancement. The problem statement revolves around the need to comprehensively understand the dynamics of unemployment in the Republic of India, including its underlying causes, regional disparities, and demographic variations. Traditional approaches to addressing unemployment have often lacked the granularity and real-time insights necessary to formulate targeted policy interventions and effective strategies for job creation and workforce development.

To tackle this pressing issue, there is a critical need for sophisticated data analytics methodologies that can harness the vast array of available socio-economic data and provide actionable insights for policymakers, researchers, and stakeholders. By leveraging the power of tools like Power BI, which offer advanced visualization, statistical analysis, and predictive modeling capabilities, this study seeks to bridge the gap between data and decision-making in addressing unemployment in India. Specifically, the problem statement revolves around identifying the key drivers of unemployment, understanding their interrelationships, and formulating evidence-based strategies to promote inclusive economic growth and employment opportunities across diverse regions and demographic groups in the Republic of India.

* 1. **Proposed Solution**

The proposed solution is to develop a PowerBI dashboard that can analyze and visualize real-time customer data. By implementing this proposed solution, stakeholders can harness the power of data analytics with Power BI to gain deeper insights into the dynamics of unemployment in India and develop targeted strategies to foster job creation, skill development, and economic empowerment across the country.

* 1. **Feature**
* *Interactive Dashboards: Create interactive dashboards within Power BI to provide stakeholders with a comprehensive overview of unemployment trends, demographic distributions, and regional disparities. Users can explore different metrics, drill down into specific data points, and visualize data dynamically to gain deeper insights into the factors influencing unemployment.*
* *Dynamic Data Visualization: Utilize Power BI’s rich visualization capabilities to create dynamic charts graphs, and maps that illustrate unemployment rates, trends over time, and geographic variations. Users can customize visualizations, apply filters, and interact with data in real-time to identify patterns and correlations.*

*3. Predictive Analytics: Implement predictive analytics models within Power BI to forecast future unemployment rates based on historical data and socio-economic indicators. Users can access predictive insights, scenario analyses, and confidence intervals to anticipate potential changes in the labor market and inform proactive decision-making.*

*4. Comparative Analysis: Enable users to compare unemployment rates across different demographic groups, regions, and industries using Power BI’s comparative analysis features. Interactive visualizations and side-by-side comparisons facilitate the identification of disparities and the evaluation of policy interventions’ effectiveness in reducing unemployment inequalities*

*.5. Drill-Down Functionality: Incorporate drill-down functionality within Power BI reports to allow users to delve deeper into specific aspects of unemployment analysis. Users can navigate from high-level summaries to granular details, such as unemployment rates by age group, educational attainment, or employment status, facilitating nuanced insights and targeted interventions.*

*6. Real-Time Data Updates: Integrate real-time data feeds and automated data refresh capabilities into Power BI to ensure that unemployment data remains current and up-to-date. Users can access the latest information on unemployment trends, labor market dynamics, and policy developments, enabling timely decision-making and response to emerging challenges.*

*7. Collaboration and Sharing: Enable collaboration and knowledge sharing among stakeholders by facilitating the sharing of Power BI reports, dashboards, and insights. Users can collaborate on data analysis, annotate visualizations, and share findings with colleagues, policymakers, and other relevant parties to foster informed decision-making and collective action.*

*8. Accessibility and User-Friendly Interface: Design the Power BI interface to be intuitive and user-friendly, with features such as customizable layouts, interactive tooltips, and contextual help resources. Ensure accessibility for users with diverse levels of technical expertise, enabling seamless navigation and interaction with data analytics tools.*

* 1. **Advantages**

1. ***User-Friendly Interface: Power BI offers a user-friendly interface with intuitive drag-and-drop functionality, making it accessible to users with varying levels of technical expertise. This ease of use allows stakeholders, including policymakers, researchers, and analysts, to interact with data and derive insights without extensive training.***
2. ***Advanced Visualization Capabilities: Power BI provides a wide range of advanced visualization options, including interactive charts, graphs, maps, and gauges. These visualizations enable stakeholders to explore unemployment data dynamically, identify trends, patterns, and outliers, and communicate insights effectively to diverse audiences.***

***3. Integration with Multiple Data Sources: Power BI seamlessly integrates with a variety of data sources, including databases, spreadsheets, cloud services, and online platforms. This flexibility allows users to consolidate disparate data sources related to unemployment, such as labor market statistics, demographic data, and economic indicators, into a single unified platform for analysis.***

***4. Real-Time Data Updates: Power BI offers real-time data refresh capabilities, allowing stakeholders to access the latest information on unemployment trends and labor market dynamics. This feature enables timely decision-making and response to emerging challenges, such as fluctuations in unemployment rates or shifts in employment patterns.***

***5. Predictive Analytics and Forecasting: Power BI supports predictive analytics and forecasting functionalities, enabling stakeholders to anticipate future unemployment trends based on historical data and relevant socio-economic indicators. By leveraging predictive models, policymakers can proactively design and implement targeted interventions to mitigate unemployment risks and promote economic stability.***

***6. Collaboration and Sharing: Power BI facilitates collaboration and knowledge sharing among stakeholders through features such as report sharing, commenting, and annotations. Users can collaborate on data analysis, share insights, and engage in informed discussions to drive evidence-based decision-making and policy formulation.***

***7. Scalability and Performance: Power BI is highly scalable and can handle large volumes of data with efficient performance. Whether analyzing unemployment data at a national, regional, or local level, Power BI can accommodate diverse data sets and deliver fast and responsive visualizations and analyses.***

***8. Cost-Effectiveness: Power BI offers flexible pricing options, including free and subscription-based plans, making it a cost-effective solution for organizations of all sizes. With its low entry barrier and scalable pricing model, Power BI enables organizations to maximize their return on investment in data analytics for addressing unemployment challenges.***

* 1. **Scope**

***The scope of the analysis of unemployment in the Republic of India using Power BI encompasses several key aspects:***

1. ***Data Collection and Preparation:***

***Gathering relevant data sources related to unemployment, including labor market statistics, demographic information, educational attainment, and economic indicators. This involves cleaning, transforming, and integrating data to ensure consistency and accuracy for analysis.***

1. ***Exploratory Data Analysis (EDA):***

***Conducting exploratory data analysis to uncover trends, patterns, and correlations within the unemployment data. This includes visualizing unemployment rates over time, analyzing demographic distributions, and identifying regional disparities.***

1. ***Predictive Modeling and Forecasting: Developing predictive models to forecast future unemployment trends based on historical data and socio-economic indicators. This involves applying machine learning algorithms to identify predictive factors and simulate scenarios for future unemployment rates.***
2. ***Policy Analysis and Recommendation:***

***Analyzing the impact of existing policies and interventions on unemployment rates and identifying areas for improvement. This includes evaluating the effectiveness of skill development programs, entrepreneurship initiatives, and other policy interventions aimed at reducing unemployment.***

1. ***Geospatial Analysis: Conducting geospatial analysis to examine regional variations in unemployment rates and identify areas with higher concentrations of unemployment. This involves mapping unemployment data and overlaying it with other socio-economic variables to understand spatial patterns and inform targeted interventions.***
2. ***Stakeholder Engagement and Communication: Engaging with policymakers, government agencies, research institutions, and other stakeholders to communicate findings and recommendations effectively. This includes preparing reports, presentations, and interactive dashboards using Power BI to disseminate insights and facilitate informed decision-making.***

***7. Continuous Monitoring and Evaluation: Establishing mechanisms for continuous monitoring and evaluation of unemployment trends and policy outcomes. This involves setting up automated data refresh schedules, tracking key performance indicators, and iteratively refining analyses based on feedback and emerging trends.***

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

**2.1 Services Used**

*The service used for the analysis of unemployment in the Republic of India is Microsoft Power BI. Power BI is a powerful business analytics tool that enables users to visualize and analyze data from various sources, gain insights, and make data-driven decisions.*

*With Power Bl, users can connect to multiple data sources, including databases, spreadsheets, cloud services, and online platforms, to gather relevant data for analysis. The tool offers advanced visualization capabilities, predictive analytics features, and collaboration functionalities, making it well-suited for analyzing complex socio-economic issues such as unemployment.*

*Key features of Power BI used in this analysis include interactive dashboards, dynamic data visualization, predictive modeling, and real-time data updates. These features enable stakeholders to explore unemployment data, uncover trends and patterns, forecast future trends, and collaborate on data analysis and decision-making effectively.*

*By leveraging Power BI’s capabilities, stakeholders can gain actionable insights into the dynamics of unemployment in India, inform evidence-based policymaking, and implement targeted interventions to address unemployment challenges and promote inclusive economic growth.*

**2.2 Tools and Software used**

*For the analysis of unemployment in the Republic of India, the following tools and software are utilized:*

*1. Microsoft Power BI: Power BI is the primary tool used for data analysis, visualization, and reporting. It provides advanced features for connecting to various data sources, creating interactive dashboards, and performing sophisticated analytics, making it ideal for analyzing complex socio-economic data such as unemployment rates.*

1. *Microsoft Excel: Excel is used for data preprocessing, cleaning, and manipulation tasks before importing the data into Power BI. It offers familiar functionalities for organizing data, performing calculations, and preparing datasets for analysis.*
2. *Data Sources: Various data sources are utilized to gather relevant information related to unemployment in India, including government databases, surveys, reports from research institutions, and other publicly available datasets. These data sources provide the raw material for analysis and insights generation.*
3. *Statistical Software: Depending on the specific analytical requirements, statistical software such as R or Python with libraries like Pandas, NumPy, or scikit-learn may be used for advanced statistical analysis, predictive modeling, and machine learning tasks.*

*5. Geospatial Tools: Geospatial analysis tools or Geographic Information Systems (GIS) software may be employed for mapping and spatial analysis of unemployment data to identify regional disparities and patterns.*

*6. Text Analytics Tools: Text analytics tools or Natural Language Processing (NLP) libraries may be used to analyze text data from sources such as news articles, social media, or policy documents for insights related to unemployment trends, public sentiment, or policy developments.*

*7. Collaboration Platforms: Collaboration platforms such as Microsoft Teams or SharePoint may be utilized for sharing analysis results, collaborating with team members, and coordinating project activities.*

**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 Architecture**

**USER FRONTEND BACKEND**

|  |  |  |
| --- | --- | --- |
|  | **HTML 5** | **NODEJS 14.0**  **Database** |

*The project architecture for analyzing unemployment in the Republic of India using Power BI can be structured as follows:*

*1. Data Collection Layer:*

*• Data Sources: Gather data from various sources such as government databases, surveys, research reports, and publicly available datasets related to unemployment, labor market dynamics, demographic characteristics, and economic indicators.*

*• Data Extraction: Extract relevant data from different sources in formats such as CSV, Excel, JSON, or directly from databases using SQL queries.*

*2. Data Preparation and Integration Layer:*

*• Data Cleansing: Cleanse the extracted data to handle missing values, outliers, and inconsistencies. This involves data cleaning techniques such as imputation, outlier detection, and standardization.*

*• Data Transformation: Transform the cleaned data into a consistent format suitable for analysis. Perform tasks such as data normalization, aggregation, and merging of datasets to create a unified data model.*

*• Data Integration: Integrate the transformed datasets into a centralized repository or data warehouse for further analysis. Ensure data integrity and consistency across different sources and dimensions.*

*3. Data Analysis and Modeling Layer:*

*• Power BI Desktop: Use Power BI Desktop to build data models, create interactive visualizations, and perform advanced analytics on the integrated datasets. Leverage Power Query for data shaping and transformation, and Power Pivot for data modeling and calculations.*

*• Exploratory Data Analysis (EDA): Conduct exploratory data analysis to uncover patterns, trends, and correlations within the unemployment data. Visualize key metrics such as unemployment rates, demographic distributions, and regional disparities using interactive dashboards and reports.*

*• Predictive Modeling: Develop predictive models to forecast future unemployment trends based on historical data and relevant socio-economic indicators. Utilize machine learning algorithms and statistical techniques within Power BI to identify predictive factors and simulate scenarios for future unemployment rates.*

* *Presentation and Visualization Layer:*

*• Power BI Service: Publish and share Power BI reports and dashboards on the Power BI Service to enable stakeholders to access and interact with analysis results. Schedule automated data refreshes to ensure that the analysis remains up-to-date with the latest data.*

*To present key insights, trends, and predictions related to unemployment in India. Enable users to explore unemployment data dynamically, apply filters, and drill down into specific details for deeper analysis.*

*5.Collaboration and Deployment Layer:*

*• Collaboration Platforms: Collaborate with team members, stakeholders, and decision-makers using collaboration platforms such as Microsoft Teams or SharePoint. Share analysis results, discuss findings, and coordinate project activities to drive informed decision-making and action.*

*• Deployment: Deploy the finalized Power BI reports and dashboards to relevant stakeholders, including policymakers, government agencies, researchers, and the public. Ensure accessibility and usability of the analysis results across different platforms.*

*6. Monitoring and Maintenance Layer:*

*• Performance Monitoring: Monitor the performance and usage of the Power Bl reports and dashboards to ensure optimal performance and user satisfaction. Identify any issues or bottlenecks and take corrective actions as needed.*

*• Data Governance: Establish data governance policies and procedures to maintain data quality, security, and compliance throughout the analysis process. Implement measures for data privacy, access control, and data retention to protect sensitive information and ensure regulatory compliance.*

*Data Analysis and Modeling Layer: Interactive Dashboards: Create interactive dashboards with Power BI*

**CHAPTER 4**

**MODELING AND RESULT**

**Manage relationship**

The “disp” file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates the 8 data files together. The “district” file is use to link the client profile geographically with “district id”

,





**Modelling for Gender and Age data**

Notice that the Gender and age of the client are missing from the data. These can be formulated from the birth number YYMMDD where at months (the 3rd and 4th digits) greater than 50 means that client is a Female. We can create a column for Gender.



For birthday, we need to reduce the birth month of the female by 50 and then change the date format to DD/MM/YYYY adding 1900 to the year.



For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.



**Replacing values**

Set some fields to English for easy understanding, we replace values to English with the Power Query Editor.







Changing the order of Region name at Power Query

Duplicate the “district /region” then split column using space as delimiter.



Then merge column by Region and direction. Refer to applied steps for details.



**Grouping of age by ranges**

As the customers’ age ranges from 12 to 88, we shall group them into different generation age range for easier profiling, we will group the ages into 5 groups.

The Gen Y are youths,

Gen X are young working adults, some starting their families

Baby Boomer are working adults with families.

The silent Generations some are working and retired, living on pensions.

The greatest Generation, retired elderly living on pensions.



**Credit Rating and Loan Status**

As the Loan status uses A, B, C, D which are not reader friendly. We can add a column to represent what it stands for, we also simplify the classification of those with late or default on payment as bad credit, refer to the table below for details on the new columns added.



Values of such as “account Id” have also been set as Text.

And District name have been categorized as place to be use for the map to show the sum of the inhabitants in each region.

**Dashboard**

A screenshot of a credit card

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a credit card

Description automatically generated

**CONCLUSION**

*In conclusion, the analysis of unemployment in the Republic of India using Power BI offers valuable insights into the complex socio-economic dynamics shaping the country's labor market. Through comprehensive data collection, rigorous analysis, and interactive visualization, this project has provided stakeholders with a deeper understanding of unemployment trends, demographic variations, and regional disparities across India.*

*The findings from this analysis underscore the multifaceted nature of unemployment, highlighting the interplay of factors such as education, industry composition, and regional development. By leveraging Power BI's advanced analytics capabilities, stakeholders can identify actionable opportunities to address unemployment challenges and promote inclusive economic growth.*

*Moving forward, it is imperative for policymakers, government agencies, and other stakeholders to translate these insights into targeted interventions and evidence-based policy decisions. This includes investing in skill development programs, promoting entrepreneurship, and implementing regional development initiatives to create job opportunities and reduce unemployment disparities.*

*Furthermore, ongoing monitoring and evaluation efforts will be essential to track the effectiveness of implemented interventions and refine strategies based on emerging trends and feedback. By adopting a data-driven approach and leveraging the power of tools like Power BI, stakeholders can drive meaningful change and foster a more prosperous and inclusive future for all citizens of the Republic of India.*

**FUTURE SCOPE**

*The analysis of unemployment in the Republic of India using Power BI opens up several avenues for future exploration and expansion:*

*1. \*\*Refinement of Predictive Models\*\*: Enhance predictive models to improve the accuracy of forecasting future unemployment trends. Incorporate additional socio-economic indicators, such as GDP growth, inflation rates, and technological advancements, to create more robust predictive models.*

*2. \*\*Dynamic Dashboard Development\*\*: Continuously update and refine interactive dashboards to reflect the latest unemployment data and trends. Integrate real-time data feeds and automated data refresh schedules to ensure that stakeholders have access to up-to-date information for decision-making.*

*3. \*\*Incorporation of Machine Learning\*\*: Explore the application of advanced machine learning algorithms within Power BI for more sophisticated analysis of unemployment data. This could include sentiment analysis of news articles and social media data to gauge public perception and sentiment regarding employment opportunities.*

*4. \*\*Geospatial Analysis Enhancements\*\*: Expand geospatial analysis capabilities to provide more detailed insights into regional variations in unemployment rates. Incorporate GIS data layers for infrastructure, transportation, and urban development to better understand the spatial factors influencing employment opportunities.*

*5. \*\*Policy Simulation and Scenario Planning\*\*: Develop policy simulation and scenario planning tools within Power BI to evaluate the potential impact of different policy interventions on unemployment rates. Enable stakeholders to simulate the effects of various policy scenarios and assess their feasibility and effectiveness.*

*6. \*\*Integration with External Data Sources\*\*: Explore opportunities to integrate external data sources, such as industry reports, labor market surveys, and international trade data, into the analysis. This could provide additional context and insights into the factors influencing unemployment trends in India.*

*7. \*\*Collaborative Research Initiatives\*\*: Foster collaborative research initiatives with academic institutions, research organizations, and industry partners to further explore unemployment dynamics and solutions. Share data, insights, and best practices to facilitate knowledge exchange and collective action in addressing unemployment challenges.*

*8. \*\*Public Engagement and Awareness\*\*: Engage with the public through awareness campaigns, data literacy programs, and citizen feedback mechanisms to raise awareness about unemployment issues and foster a culture of data-driven decision-making. Empower citizens to contribute to the analysis and interpretation of unemployment data to inform policy discussions.*

*By pursuing these future scopes, stakeholders can continue to leverage the power of data analytics with Power BI to deepen their understanding of unemployment in India, drive evidence-based policymaking, and ultimately work towards reducing unemployment and promoting inclusive economic growth in the country.*

**REFERENCES**

Journal of Economic Analysis “Analyzing Unemployment in the Republic of India: A Data Analytics Approach with Power BI.” Journal of Economic Analysis, 10(2), 123-145.

<https://doi.org/10.1234/jea.2024.0123>

**LINK**