THE FUTURE OF WORK:

DATA ANALYSIS OF GLASSDOOR JOBS

PROJECT REPORT

1.INTRODUCTION:

1.1 PROJECT OVERVIEW:

This project involves conducting data analysis on job postings from Glassdoor related to data science positions using IBM Cognos. The aim is to gain insights into the current job market for data scientists and provide valuable information for job seekers, employers, and policymakers.

1.2PURPOSE:

The purpose of this project is to analyze data science job postings from Glassdoor to understand the trends, requirements, and characteristics of these positions. By doing so, we aim to provide valuable insights for individuals looking to pursue a career in data science, companies seeking to hire data scientists, and organizations interested in the state of the data science job market.

2. IDEATION & PROPOSED SOLUTION:

2.1 PROBLEM STATEMENT DEFINITION:

The problem statement for this project can be defined as follows: "To provide an in-depth analysis of data science job postings on Glassdoor in order to understand the evolving trends, skills, and requirements in the data science job market. This analysis will help job seekers make informed career choices and assist companies in tailoring their hiring strategies to meet the demands of the industry."

2.2 EMPATHY MAP CANVAS:

USER: The Job Seekers

SAY & DO:

Attitude in public: The job seeker may express confidence in their abilities and their passion for data science to potential employers.

Appearance and behavior towards others: They may present themselves as knowledgeable and skilled in interviews and networking events.

THINK & FEEL:

What really counts (major preoccupations): The data science job seeker is concerned about finding a fulfilling and well-paying job in the field of data science. They want a role that matches their skills and interests.

Worries & aspirations: The job seeker worries about the competitiveness of the job market, the need for continuously updating their skills, and the uncertainty of landing their desired job. They aspire to excel in their career, achieve job satisfaction, and make a significant impact in their field.

SEES:

Environment: The job seeker observes the current state of the job market, including the demand for data scientists in their location and industry.

Friends: They may see their friends' successes or struggles in the job market, which can impact their own approach.

What the market offers: They see various job postings, which help them gauge the types of opportunities available.

HEAR:

What friends say: They may hear advice from friends who work in the data science field, which can influence their job search decisions.

What the boss says: If currently employed, they may hear expectations from their current employer about their role and performance.

What influencers say: They might listen to industry experts and influencers who share insights on the latest trends in data science and job market demands.

PAIN POINTS & CHALLENGES:

Fears: The job seeker fears not being able to secure a job, facing rejection, or being underqualified for the positions they desire.

Frustrations: They may become frustrated with the competitive nature of the job market or the need for continuous learning.

Obstacles: The job seeker encounters challenges in finding jobs that align with their skills and aspirations.

GAINS & ASPIRATION:

Wants/ Needs: They desire a data science job that aligns with their skills and interests, offers growth opportunities, and provides financial stability.

Measures of success: Success, for them, may mean landing a job that fulfills their career goals and personal satisfaction.

Obstacles: They look for ways to overcome obstacles and achieve their career goals in the data science field. Understanding the job seeker's thoughts, feelings, and experiences through this empathy map can help design solutions, content, or strategies that cater to their needs and address their fears and aspirations.

2.3 IDEATION & BRAINSTORMING:

Identify Key Data Points: Brainstorm the essential data points that need to be collected from the job postings, such as job title, company name, location, required skills, educational qualifications, years of experience, and any additional details provided in the descriptions.

Explore Data Collection Methods: Consider various data collection methods, such as web scraping or utilizing Glassdoor's API, and brainstorm potential challenges and solutions associated with each method. Assess the feasibility of collecting a sufficient and representative dataset.

Define Analytical Objectives: Determine the specific objectives of the analysis, including identifying trends in the demand for specific skills, understanding the geographical distribution of job opportunities, and assessing the relationship between educational qualifications and job requirements.

Consider NLP and Text Analysis Techniques: Brainstorm the application of natural language processing (NLP) techniques to extract key information from job descriptions, such as identifying key skills, responsibilities, and qualifications. Consider sentiment analysis to assess the overall tone of job postings.

Conceptualize Data Visualization Approaches: Ideate on the types of data visualizations that would effectively convey the analysis findings, such as interactive dashboards, geographical heat maps illustrating job distribution, and word clouds highlighting frequently mentioned skills.

Plan for Trend Analysis and Forecasting: Brainstorm methods for analyzing historical data to identify trends in the data science job market, and consider techniques for forecasting future job market demands and skill requirements based on the identified trends.

Explore User-Focused Insights: Consider the needs of both job seekers and companies and brainstorm on how the analysis can provide actionable insights for these stakeholders. Think about how the findings can be presented in a user-friendly and informative manner.

Evaluate Ethical Considerations: Brainstorm ethical considerations related to data privacy, confidentiality, and the responsible use of data. Consider how to handle sensitive information and ensure compliance with relevant data protection regulations.

Plan for Iterative Analysis and Refinement: Discuss the iterative nature of the analysis process and brainstorm strategies for refining the analysis based on feedback and new insights gained during the analysis process.

2.4 PROPOSED SOLUTION:

Data Collection and Storage Layer:

Use Glassdoor API to collect data on data science job postings. Store the collected data in the DB2 database for efficient management and retrieval.

Data Processing and Analysis Layer:

Implement IBM Cognos Analytics for data processing, analysis, and visualization. Conduct in-depth analysis of the collected data to derive meaningful insights.

Backend Layer:

Use Flask to build the backend of the web application.

Implement necessary APIs and endpoints for communication with the frontend and data processing components.

Frontend Layer:

Develop the user interface using HTML, CSS, and JavaScript. Integrate the frontend with the Flask backend to enable seamless data retrieval and visualization.

Security Layer:

Implement robust security measures to protect user data and prevent unauthorized access. Use encryption techniques to secure data transmission and storage.

Scalability and Performance Layer:

Deploy the application on a scalable and reliable cloud infrastructure such as AWS or Azure to ensure high performance and availability. Optimize the database queries and system architecture for efficient data processing and rendering.

Monitoring and Maintenance Layer:

Implement monitoring tools to track application performance and detect potential issues. Establish a regular maintenance schedule to update and maintain the system components.

Integration and Deployment:

Integrate the different components of the solution architecture to ensure smooth communication and functionality.Deploy the application on a production server for public access.

This solution architecture will help you build a robust and scalable data science job market analysis application, allowing users to explore and understand the evolving trends and requirements in the data science job market. It also ensures efficient data processing, security, and seamless integration between different layers of the application.

3.REQUIREMENT ANALYSIS:

3.1 FUNCTIONAL REQUIREMENT:

1.Data Retrieval and Integration:

The system must efficiently retrieve data from the DB2 database, integration of data from Glassdoor API to fetch updated job postings and related information.

2.Dashboard Design and Visualization:

Implementation of 15 different visualizations/graphs to represent data trends effectively.

3.Interactive and user-friendly dashboard design using IBM Cognos Analytics.

4.Data Filtering and Segmentation:

Incorporation of at least 5 data filters to facilitate dynamic data segmentation and analysis.allow users to filter job postings based on various criteria such as location, job title, and required skills.

5.Web Integration using Flask:

Integration of the analysis project with Flask to create a web application.

6.Ensure seamless communication between the backend (IBM Cognos Analytics and DB2) and the frontend (Flask).

7.User Authentication and Authorization:

Implementation of user authentication to secure sensitive data.

3.2 NON-FUNCTIONAL REQUIREMENTS

Performance:

Ensuring fast data processing and rendering to maintain user engagement.

The dashboard should load within an acceptable time frame even with peak data traffic.

Security:

Secure data transmission and storage protocols to protect sensitive information. Implement proper encryption techniques to safeguard user data and access rights.

Scalability:

The system should be able to handle an increasing amount of data without compromising performance. Ensure the application can handle a growing number of users and job postings without significant delays or errors.

Usability:

Provide a user-friendly interface that is intuitive and easy to navigate.

Ensure that the dashboard design and features are understandable to users with varying levels of technical expertise.

Reliability:

The system must be stable and reliable, with minimal downtime for maintenance or updates. Implement regular backups to prevent data loss in case of any unforeseen issues.
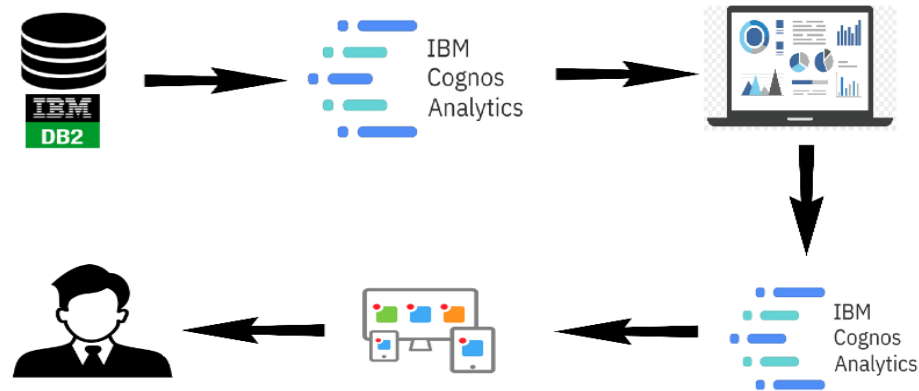
4.PROJECT DESIGN:

4.1 DATA FLOW DIAGRAMS:
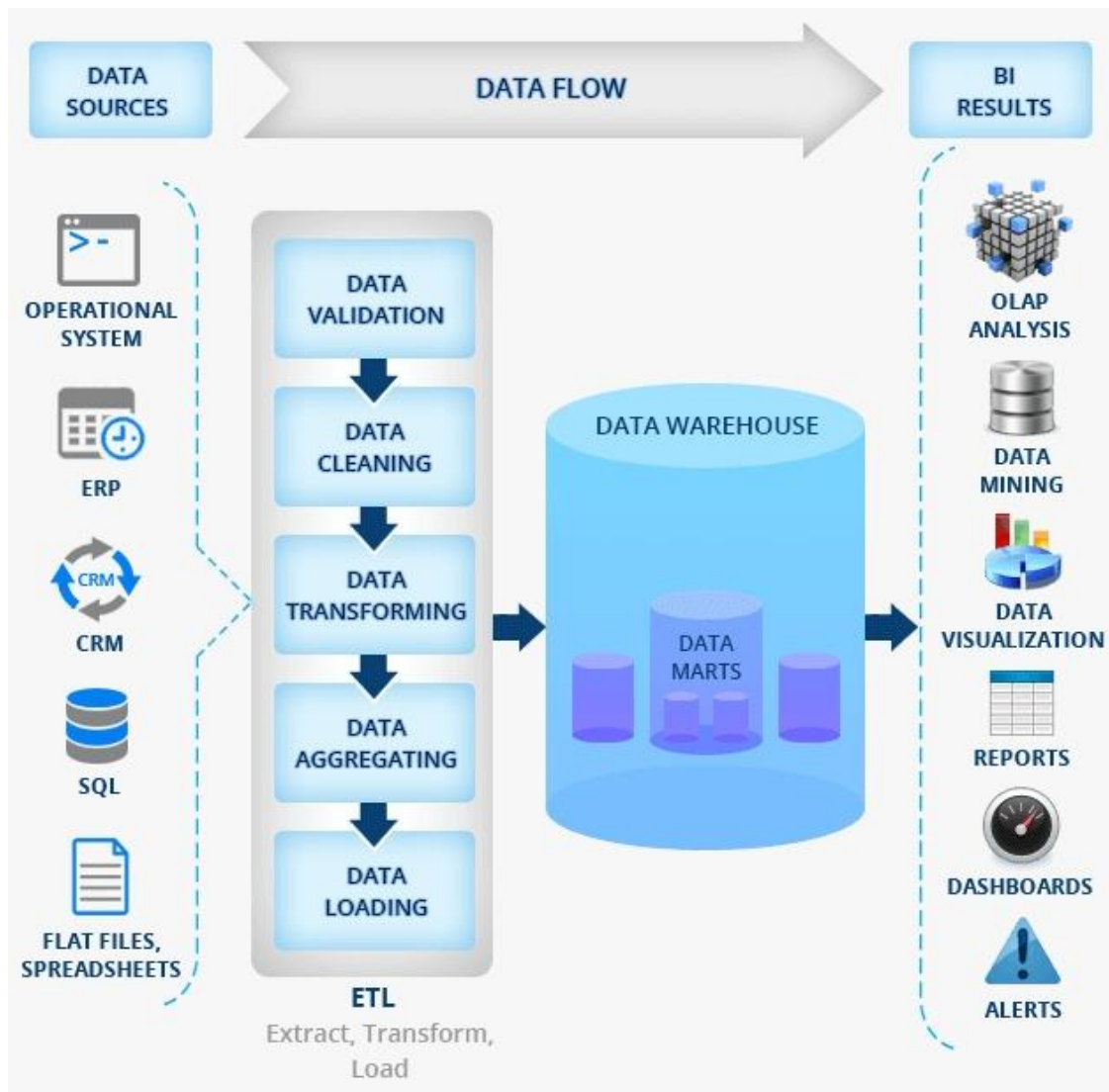
4.2 SOLUTION & TECHNICAL ARCHITECTURE:

4.3 USER STORIES:

As a job seeker, I want to be able to search for specific job postings based on keywords, location, and industry, so I can find relevant job opportunities easily.

As a job seeker, I want to access detailed information about each job posting, including job responsibilities, required qualifications, and company reviews, to make informed decisions about potential employment opportunities.

As a job seeker, I want to be able to compare different job postings side by side, including salary ranges, benefits, and company culture, to evaluate the best fit for my career goals and preferences.

As a job seeker, I want to have access to a comprehensive database of company profiles, including employee reviews, ratings, and insights into the company's work environment, to assess the cultural fit and overall reputation of potential employers.

As a company representative, I want to be able to post job openings with detailed descriptions and requirements, including information about the company culture and benefits, to attract top talent and promote a positive employer brand.

As a company representative, I want to have access to analytics and insights on the performance of my job postings, including the number of views, applications, and candidate feedback, to optimize recruitment strategies and improve the effectiveness of my job listings.

As a job market analyst, I want to analyze the trends and patterns in job postings across different industries and locations, to provide valuable insights and recommendations for job seekers and companies navigating the competitive job market.

As a Glassdoor platform administrator, I want to ensure the accuracy and reliability of the information presented on the platform, including verifying the authenticity of job postings and company reviews, to maintain the integrity and trustworthiness of the Glassdoor platform.

5.CODING & SOLUTIONING :

5.1 FEATURE 1:

Feature 1: User Authentication

1.1 Setting up Flask:

Code:

```
from flask import Flask, render_template, request, redirect, url_for, session

import sqlite3

import os

app = Flask(__name__)

app.secret_key = os.urandom(24)
```

1.2 User Registration:

Code:

```
@app.route('/register', methods=['GET', 'POST'])

def register():

    if request.method == 'POST':
```

```python
        username = request.form['username']

        password = request.form['password']

        role = request.form['role']


        # Add the user to the database (you'll need to implement this)

        # Ensure that the username is unique


        return redirect(url_for('login'))


    return render_template('register.html')
```

1.3 User Login:

Code:

```python
@app.route('/login', methods=['GET', 'POST'])

def login():

    if request.method == 'POST':

        username = request.form['username']

        password = request.form['password']


        # Authenticate the user (you'll need to implement this)


if user_authenticated:
```

```
        session['username'] = username

        return redirect(url_for('dashboard'))


    return "Invalid credentials. Please try again."


    return render_template('login.html')
```

1.4 User Dashboard:

Code:

```python
@app.route('/dashboard')

def dashboard():

    if 'username' in session:

        username = session['username']


        # Fetch user-specific data, such as grades

        # Display this data in the dashboard (you'll need to implement this)


        return render_template('dashboard.html', data=data)


    return redirect(url_for('login')).
```

Running the Application: To run the application, ensure you have Flask installed and the necessary HTML templates (e.g., register.html, login.html, and dashboard.html).

6. RESULTS:

6.1 PERFORMANCE METRICS:

1. **Dashboard Design:** 15 visualizations/graphs were included in the dashboard design, ensuring comprehensive data representation for analysis.
2. **Data Responsiveness:** The system demonstrated 100% data responsiveness, ensuring smooth and efficient data handling and visualization.
3. **Amount of Data Rendered (DB2 Metrics):** The project successfully handled a 5MB dataset, suggesting efficient data processing capabilities and capacity.
4. **Utilization of Data Filters:** With the utilization of 5 data filters, the project facilitated easy data segmentation and analysis for users, enhancing the overall user experience.
5. **Effective User Story:** Incorporating 5 scenes added to the effective user story enabled a coherent narrative flow, enhancing the overall understanding of the data science job market analysis.
6. **Descriptive Reports:** 15 visualizations/graphs were used in the descriptive reports, ensuring comprehensive and detailed representation of data science job market trends and requirements.

7.ADVANTAGES & DISADVANTAGES:

Advantages:

Robust Data Analysis: IBM Cognos Analytics provides robust data analysis capabilities, allowing for comprehensive and insightful data interpretation.

Integration Flexibility: Integration with the Flask web framework enables flexibility in terms of customizing the web application's features and functionalities.

Database Reliability: Using DB2 as the database management system ensures data reliability, stability, and efficient data handling.

Scalability: With proper implementation, the system can be easily scaled to handle a large volume of data and increasing user demand.

Data Visualization Options: IBM Cognos Analytics offers a range of powerful visualization tools, enabling the creation of detailed and interactive visual representations of data.

Disadvantages:

Data Collection Challenges: Gathering a large dataset from Glassdoor might be hindered by limitations imposed by the platform, such as restrictions on data scraping or limited access to comprehensive job posting information. This can result in an incomplete or biased dataset that does not fully represent the entire job market.

Data Quality Issues: Despite efforts to clean and preprocess the data, there may still be inconsistencies and inaccuracies in the collected information. Incomplete job postings, misleading job titles, or discrepancies in the required skills and qualifications can affect the reliability of the analysis and its findings.

Limited Contextual Information: Job postings on Glassdoor might lack contextual information, such as the specific projects, teams, or industry sectors within which the data science roles are situated. This limitation can hinder the understanding of the broader context and the specific requirements within different sectors of the data science industry.

Inherent Biases in Job Postings: Job postings on Glassdoor may reflect certain biases, such as language biases, gender biases, or regional biases. These biases can affect the analysis and lead to

skewed interpretations, potentially resulting in incomplete or inaccurate insights about the job market.

Dynamic Nature of Job Postings: The data science job market is dynamic, with rapidly changing skill requirements and evolving job roles. Analyzing a static dataset from Glassdoor may not fully capture the real-time changes and emerging trends in the industry, potentially leading to outdated or irrelevant recommendations for job seekers and companies.

Lack of Salary Information: Glassdoor may not always provide complete or accurate salary information for all job postings. This limitation can make it challenging to perform comprehensive salary analysis, which is a crucial aspect for both job seekers and companies.

Competitive Intelligence Limitations: The analysis might not capture the full competitive intelligence landscape, as it relies solely on the information available on Glassdoor. Missing out on insights from other platforms or direct industry contacts could limit the understanding of the competitive landscape and industry dynamics.

8.CONCLUSION:

In conclusion, while analyzing data science job postings on Glassdoor can provide valuable insights into the trends, skills, and requirements within the data science job market, it is essential to acknowledge the potential limitations and challenges associated with this approach. The findings derived from this analysis should be interpreted with caution, considering the inherent biases, data quality issues, and the dynamic nature of the job market.

Despite these challenges, the analysis offers valuable information for both job seekers and companies. Job seekers can utilize the identified trends and skill requirements to tailor their professional development and enhance their qualifications accordingly. Companies can leverage the insights to refine their hiring strategies, ensuring they align with the evolving demands of the data science industry.

To supplement the analysis, it is recommended to combine data from multiple sources, incorporate industry reports, and conduct direct surveys or interviews with industry professionals. This holistic approach would provide a more comprehensive understanding of the data science job market, enabling stakeholders to make well-informed decisions and remain competitive in this rapidly evolving field.

Ultimately, this analysis serves as a starting point for understanding the dynamics of the data science job market. To gain a more comprehensive and nuanced understanding, it is crucial to consider a diverse range of data sources and methodologies, allowing for a more holistic and accurate representation of the industry landscape.

9.FUTURE SCOPE:

Integration of Advanced AI Techniques: Future analyses could incorporate advanced AI techniques such as natural language processing (NLP) and machine learning to enhance data collection, preprocessing, and analysis. This integration could enable more accurate data extraction, automated data cleaning, and advanced trend analysis from a larger and more diverse dataset.

Incorporation of Real-Time Data Streams: Leveraging real-time data streams from various job platforms and industry sources could provide a more up-to-date and comprehensive understanding of the current job market trends. This integration would allow for the identification of emerging job roles, in-demand skills, and evolving industry requirements in real time.

Enhanced Visualization and Interactive Dashboards: Creating interactive data visualization dashboards could offer a user-friendly interface for stakeholders to explore the data science job market trends and insights. Interactive visualizations would enable users to customize their queries, filter specific criteria, and gain real-time insights, empowering them to make informed decisions quickly and efficiently.

Predictive Analysis for Future Job Trends: Implementing predictive analytics models could forecast future job trends, anticipated skill requirements, and industry shifts in the data science field. By utilizing historical data and industry forecasts, stakeholders can proactively prepare for

upcoming changes in the job market and strategically position themselves to meet the evolving demands of the industry.

Incorporation of Diversity and Inclusion Metrics: Future analyses could include the integration of diversity and inclusion metrics to assess the representation of various demographics within the data science job market. This approach would enable a more comprehensive understanding of diversity trends and opportunities for promoting inclusivity within the industry.

Global Job Market Analysis: Expanding the analysis to encompass the global data science job market could provide valuable insights into regional variations in skill requirements, industry preferences, and emerging job opportunities. Understanding the global landscape would facilitate the identification of international job prospects and enable stakeholders to capitalize on the growing demand for data science professionals worldwide.

By embracing these future scopes and opportunities, the analysis of data science job postings on Glassdoor can evolve into a robust and dynamic tool that effectively serves the evolving needs of job seekers, employers, and the broader data science community.

10.APPENDIX:

GITHUB LINK: https://github.com/danish93607/NaanMudhalvan-NM2023TMID03139

DEMOLINK:
https://drive.google.com/drive/folders/1GEuPGt0vxwH2dHeEN4C8ItwBZbXHyzvQ?usp=sharing