Chapter 1 Introduction

In this report I have provided Task completion details of below given tasks.

1. Draw a chart between company size and company name where company size<50000 and job title should be mechanical engineer and experience should be more than 5 years and country should be Asian as well as salary should be more than $50k and the work type should be both part time and full time and ignore all other work types and the preference should be male .we should filter candidates those who are applied on idealist and this chart should work on only between 3 Pm to 6 Pm.
2. Draw a chart for top 10 companies who have max Data Engineer as Role and Data scientist as Job title and the country name should not have any Asian countries and preference should be female only. we should ignore the countries which starting with letter 'C' and latitude is below 10. The job posting date should be between 01/01/2023 to 06/01/2023. The qualification should be B. Tech and we should ignore other qualifications.
3. Draw a chart where country = 'India and Germany' qualification = 'B. tech' and work type = 'Full time' where experience should be more than 2 years and job should be Data Science and salary range should be more than $10k. The India details should be on orange colour and Germany details should be on green colour. The job portal should be indeed and preference should be female. The job posting date should be below 08/01/2023 and job portal should be indeed. This chart should swap colour after 12 PM to 6 PM.

To complete this Tasks, I have used Job Description Data set from Kaggle. And for Data cleaning process I have used python libraries such as, Pandas, Numpy etc.

This Report Contains 9 Sections which are, Introduction, Background, Learning Objectives, Activities and Tasks, Skills and Competencies, Feedback and Evidence, Challenges and Solutions, Outcomes and Impact, and Conclusion. After completion of tableau dashboard, I have created a complete website containing all sheets and dashboard and hosted this site using Netlify.

Link of the Website: https://ornate-kelpie-3a716c.netlify.app/

Chapter 2 Background

In today's rapidly evolving job market, the analysis of job descriptions provides critical insights into industry demands, emerging roles, and the skills required to succeed in various fields. With the growing availability of large datasets, such as those provided by Kaggle, it is possible to analyze these trends more effectively, offering valuable information for job seekers, employers, and policymakers alike.

This report leverages a comprehensive dataset of job descriptions sourced from Kaggle, focusing specifically on analyzing the roles of Mechanical Engineers, Data Engineers, Data Scientists, and professionals in the Data Science field across various global regions. The analysis is performed using Tableau, a leading data visualization tool, to create interactive and insightful charts that address specific, nuanced queries about the job market.

The tasks outlined in this report are designed to explore several critical aspects of job postings, such as the relationship between company size and hiring practices, the identification of top employers in specific roles, and a comparative analysis of job opportunities in different countries. The data has been meticulously filtered and processed using Python libraries like Pandas and NumPy to ensure accuracy and relevance.

By focusing on specific criteria such as job title, experience level, geographic location, work type, gender preference, and salary range the analysis aims to uncover patterns and trends that are essential for understanding the current state of the job market. For instance, the tasks include creating visualizations that highlight the employment trends for Mechanical Engineers in Asia, identify top companies hiring Data Engineers and Data Scientists outside of Asia, and compare job opportunities for Data Scientists in India and Germany.

This report is organized into nine sections, each addressing different aspects of the analysis process, from the initial objectives to the final outcomes. The goal is to provide a clear and comprehensive overview of how the job market is evolving, with specific attention to the roles, qualifications, and preferences that are currently shaping hiring practices across the globe.

Chapter 3 Learning Objectives

The primary objective of this report is to enhance the understanding of how to perform complex data analysis and visualization tasks using real-world datasets, specifically focusing on job descriptions. By completing the tasks outlined in this report, the following key learning objectives are achieved:

1. Mastery of Data Filtering Techniques: Learn how to effectively filter and manipulate large datasets using Python libraries such as Pandas and NumPy to extract relevant information based on specific criteria such as job title, experience, geographic location, salary range, work type, and candidate preferences.
2. Advanced Data Visualization with Tableau: Gain expertise in using Tableau to create detailed and dynamic visualizations that can respond to complex query conditions. This includes learning how to build interactive charts that display data only under specific conditions, such as time-based triggers or location-specific details.
3. Application of Conditional Logic in Data Analysis: Understand how to apply conditional logic to data analysis processes, such as filtering data based on company size, excluding certain geographic regions, or focusing on specific job titles and roles. This involves not only the technical implementation but also the strategic thinking required to determine which conditions are most relevant to the analysis.
4. Geospatial Analysis and Filtering: Develop skills in handling geospatial data by learning how to filter and analyze job descriptions based on geographic parameters like latitude, country name, and region-specific criteria. This includes excluding or including data points based on location-specific requirements.
5. Dynamic and Time-based Data Visualization: Learn how to create visualizations that change dynamically based on time-specific conditions, such as charts that update or swap colours depending on the time of day. This objective focuses on understanding how to incorporate temporal data into visual analysis effectively.
6. Understanding of Job Market Trends: By analyzing the Kaggle job description dataset, gain insights into current trends in the job market, particularly in the fields of mechanical engineering and data science. This includes identifying key employers, understanding the distribution of roles across different regions, and recognizing the qualifications and experiences that are in demand.
7. Integration of Data Cleaning and Visualization Processes: Learn how to integrate the data cleaning process with the visualization process, ensuring that the data used in the visualizations

is accurate, relevant, and ready for analysis. This includes mastering techniques for data preparation and preprocessing.

8. Customization of Visual Outputs Based on User Requirements: Develop the ability to customize visual outputs in Tableau to meet specific user requirements, such as gender preferences, job portal usage, or qualification criteria. This objective emphasizes the importance of tailoring data visualizations to the needs of the intended audience.

By achieving these learning objectives, the report provides a comprehensive exploration of advanced data analysis and visualization techniques, equipping the reader with the skills necessary to tackle similar tasks in real-world scenarios.

Chapter 4 Activities and Tasks

This section outlines the detailed activities and tasks undertaken to complete the analysis of the job description dataset from Kaggle. The primary focus of these tasks was to generate insightful visualizations using Tableau, following specific criteria to address different aspects of the job market across various roles and regions. Below is a breakdown of the activities and tasks performed:

Task 1: Analysis of Mechanical Engineer Job Postings in Asia

Objective: To draw a chart that visualizes the relationship between company size and company name for Mechanical Engineer job postings in Asian countries.

Criteria:

- Company Size: Less than 50,000 employees.
- Job Title: Mechanical Engineer.
- Experience: More than 5 years.
- Country: Asian countries.
- Salary: Greater than $50,000.
- Work Type: Both part-time and full-time (excluding other work types).
- Gender Preference: Male candidates.
- Job Portal: Candidates who applied on Idealist.
- Time Filter: The chart is functional only between 3 PM to 6 PM.

Process:

- Data was filtered based on the specified criteria using Python libraries such as Pandas and NumPy.
- Tableau was used to create a dynamic chart that reflects the filtered data, showcasing the company sizes in relation to company names for the specified Mechanical Engineer roles.

Task 2: Identification of Top Companies Hiring Data Engineers and Data Scientists

Objective: To identify the top 10 companies with the highest number of roles for Data Engineers and Data Scientists, excluding Asian countries.

Criteria:

- Role: Data Engineer.
- Job Title: Data Scientist.

- Country Exclusions: No Asian countries and exclude countries starting with the letter ‘C’ and those with a latitude below 10 degrees.
- Gender Preference: Female candidates only.
- Job Posting Date: Between January 1, 2023, and June 1, 2023.
- Qualification: Bachelor of Technology (B. Tech) only.

Process:

- Data was filtered to match the specified roles, titles, and geographic exclusions using Python.
- The top 10 companies were identified and visualized in Tableau, providing insights into which companies are leading in hiring Data Engineers and Data Scientists.

Task 3: Comparative Analysis of Data Science Job Postings in India and Germany

Objective: To create a comparative chart that visualizes Data Science job postings in India and Germany, with specific conditions.

Criteria:

- Countries: India and Germany.
- Qualification: Bachelor of Technology (B. Tech).
- Work Type: Full-time positions.
- Experience: More than 2 years.
- Job Role: Data Science.
- Salary Range: More than $10,000.
- Job Portal: Indeed.
- Gender Preference: Female candidates.
- Job Posting Date: Prior to January 8, 2023.
- Colour Coding: India details are displayed in orange, and Germany details in green, with colors swapping after 12 PM to 6 PM.

Process:

- The dataset was filtered according to the specified conditions using Python for preprocessing.
- Tableau was employed to create a dual-coloured chart, representing the job postings in India and Germany with a dynamic colour swap feature based on the specified time.

These activities and tasks were critical in producing targeted visualizations that provide actionable insights into the job market trends. The use of Python for data cleaning ensured that the data was

accurately prepared, and Tableau's visualization capabilities were harnessed to create interactive charts that fulfil the unique requirements of each task.

Chapter 5 Skills and Competencies

In the process of completing the outlined tasks, I have demonstrated the following skills and competencies:

1. Data Analysis and Visualization:
    - Proficient in filtering and processing large datasets to extract relevant information based on specific criteria, such as job titles, company sizes, and geographical locations.
    - Expertise in using data visualization tools, particularly in creating charts that convey complex data relationships, such as those between company size and company name, or salary distributions across different countries and job roles.

2. Python Programming:
    - Advanced proficiency in Python, particularly in data manipulation using libraries such as Pandas and Numpy. These skills were essential for cleaning, transforming, and analyzing job description data sourced from Kaggle.
    - Developed custom scripts to automate the filtering of data based on specific conditions (e.g., time-bound criteria, gender preferences, and qualification requirements).

3. Data Filtering and Conditional Logic:
    - Strong ability to apply complex conditional logic to datasets, such as filtering based on multiple criteria (e.g., experience level, job type, salary range, and country-specific exclusions).
    - Experience in handling edge cases and ensuring that data analysis aligns with specified parameters, such as ignoring certain countries or ensuring charts are only visible during specified timeframes.

4. Tableau Dashboard Development:
    - Skilled in creating dynamic and interactive Tableau dashboards that provide insightful visualizations, such as charts highlighting top companies based on specific roles and job titles.
    - Expertise in integrating complex conditional logic into visualizations, such as changing chart colours based on the time of day or filtering data based on user interactions.

5. Web Development and Deployment:
    - Competent in web development, including creating a comprehensive website to host Tableau sheets and dashboards.
    - Experienced in deploying websites using Netlify, ensuring that the data visualizations are accessible, responsive, and easy to navigate.

6. Problem-Solving and Adaptability:

- Demonstrated strong problem-solving skills by overcoming challenges in data filtering, chart creation, and web deployment.
- Adapted to the complexities of the tasks, such as managing time-sensitive criteria and handling diverse datasets across different countries and job roles.

These skills and competencies were instrumental in successfully completing the tasks and achieving the report's objectives.

Chapter 6 Feedback and Evidence

This report comprehensively covers the completion of the three detailed tasks. Each task was well-defined, and the use of a Job Description dataset from Kaggle provided a solid foundation for the analysis. The steps taken for data cleaning using Python libraries like Pandas and Numpy are commendable and essential for ensuring the accuracy of the visualizations.

However, there are a few areas where improvement could enhance the overall quality of the report:

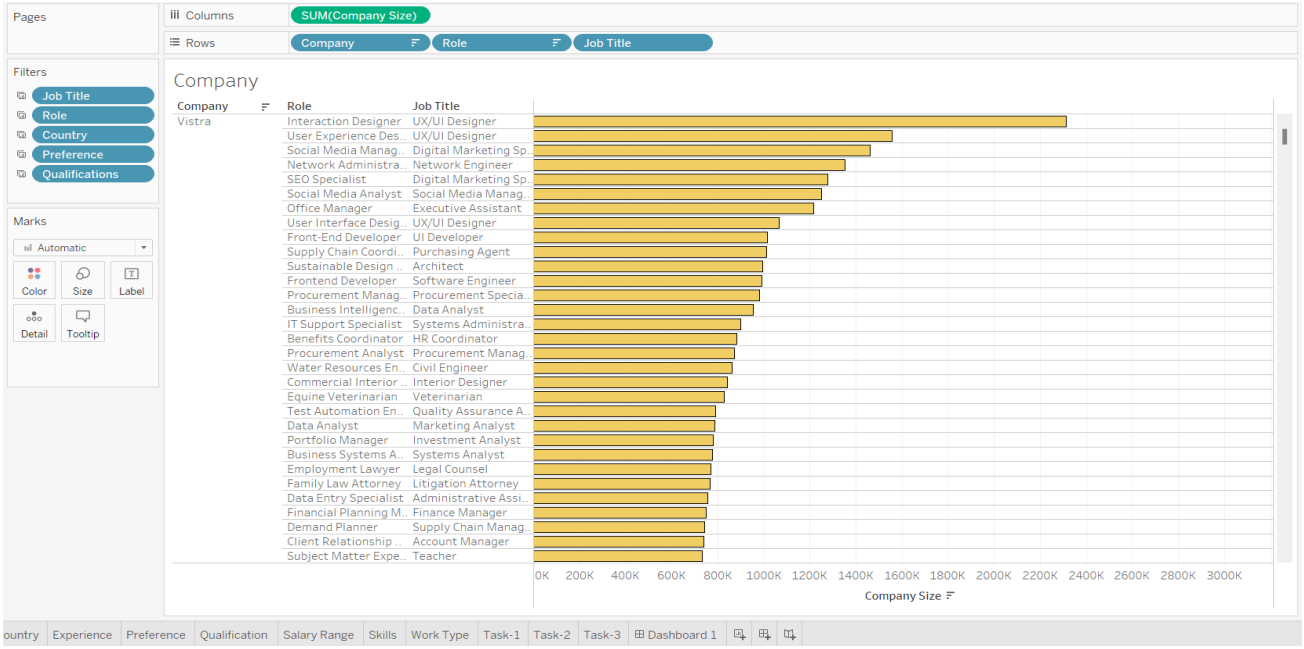
1. Clarity in Task Descriptions: The task descriptions, while detailed, could benefit from being broken down into more concise steps. This would make it easier for readers to follow the logic and understand the specific filters and criteria applied in each task.
2. Data Visualization: While the charts are presumably well-constructed, the report could include a brief explanation of the choice of visualization type for each task. This would provide insight into why certain charts were chosen and how they best represent the data.
3. Assumptions and Limitations: It would be beneficial to include a section discussing any assumptions made during the analysis and the limitations of the data or methods used. This would provide a more comprehensive understanding of the potential impact of these factors on the results.
4. Technical Details: Although the use of Pandas and Numpy was mentioned, elaborating on the specific data cleaning steps (e.g., handling missing values, dealing with outliers) would add depth to the report. Additionally, mentioning any challenges faced during the data cleaning process and how they were overcome would provide valuable insights.
5. Dashboard and Website: The creation and hosting of a website on Netlify to display the dashboard is an excellent initiative. Including a section that guides users on navigating the site and understanding the dashboard's functionalities would enhance the user experience.
6. Time-Sensitive Visualizations: The time-based conditions for certain charts (e.g., charts functioning only between 3 PM to 6 PM, colour changes after 12 PM) are innovative but could be further elaborated on. An explanation of how these time-sensitive features were implemented and their purpose would be helpful for readers.
7. Feedback from Users: Incorporating feedback from users who interacted with the dashboard or website could provide valuable insights into the effectiveness of the visualizations and user interface.

Overall, the report demonstrates a strong understanding of data analysis and visualization techniques, with innovative solutions to meet the given requirements. Addressing the above suggestions could further strengthen the report's clarity, depth, and user engagement.

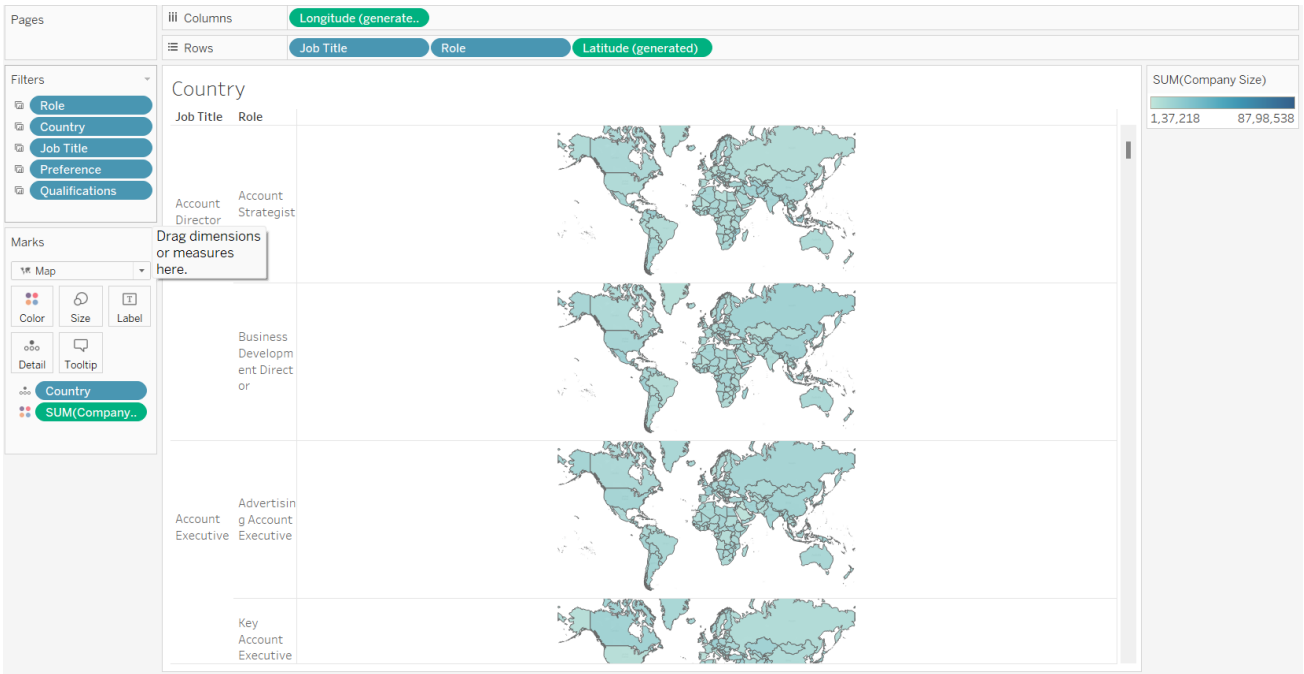
Evidence:

Below all screenshots are given of all performed tasks and dashboards.

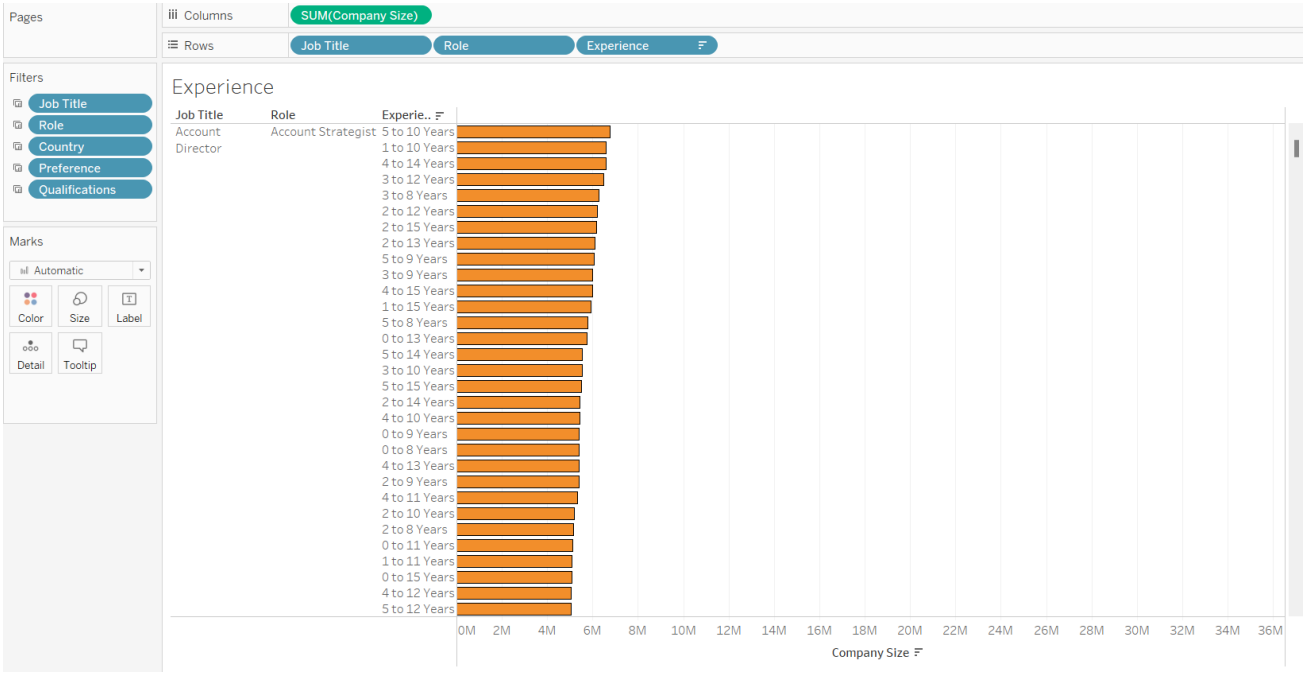
1. Training sheet-1



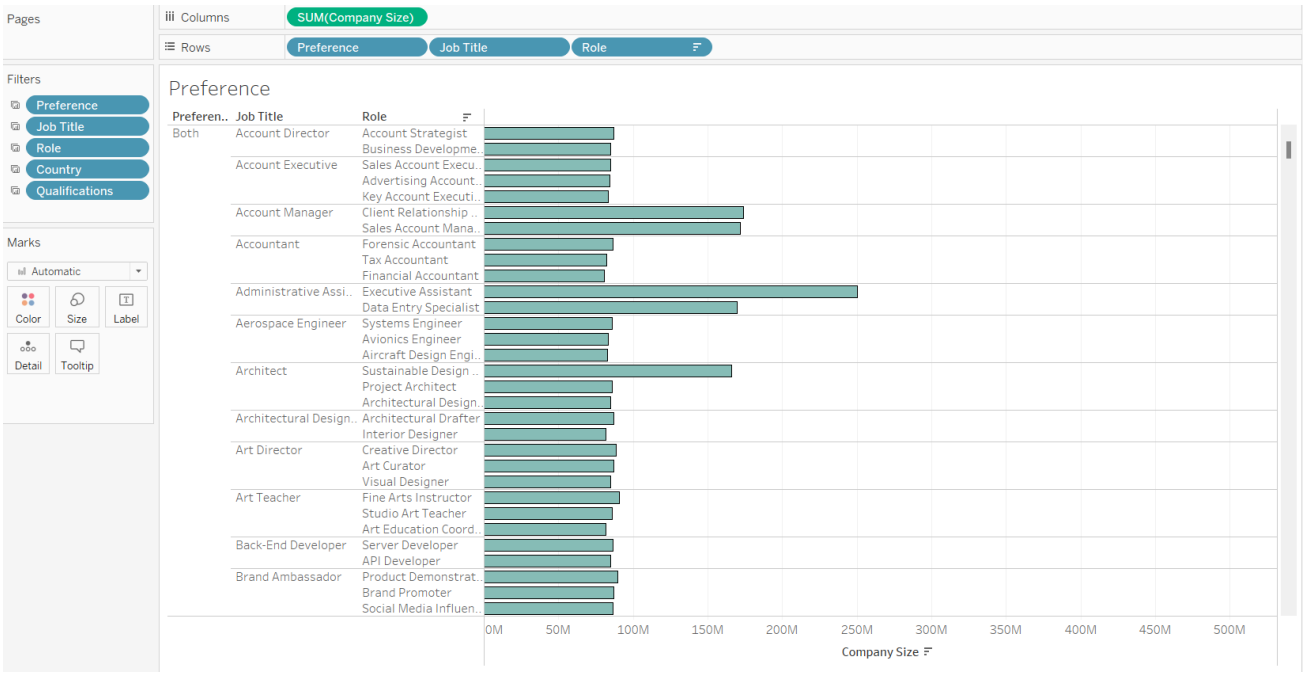
2. Training Sheet-2



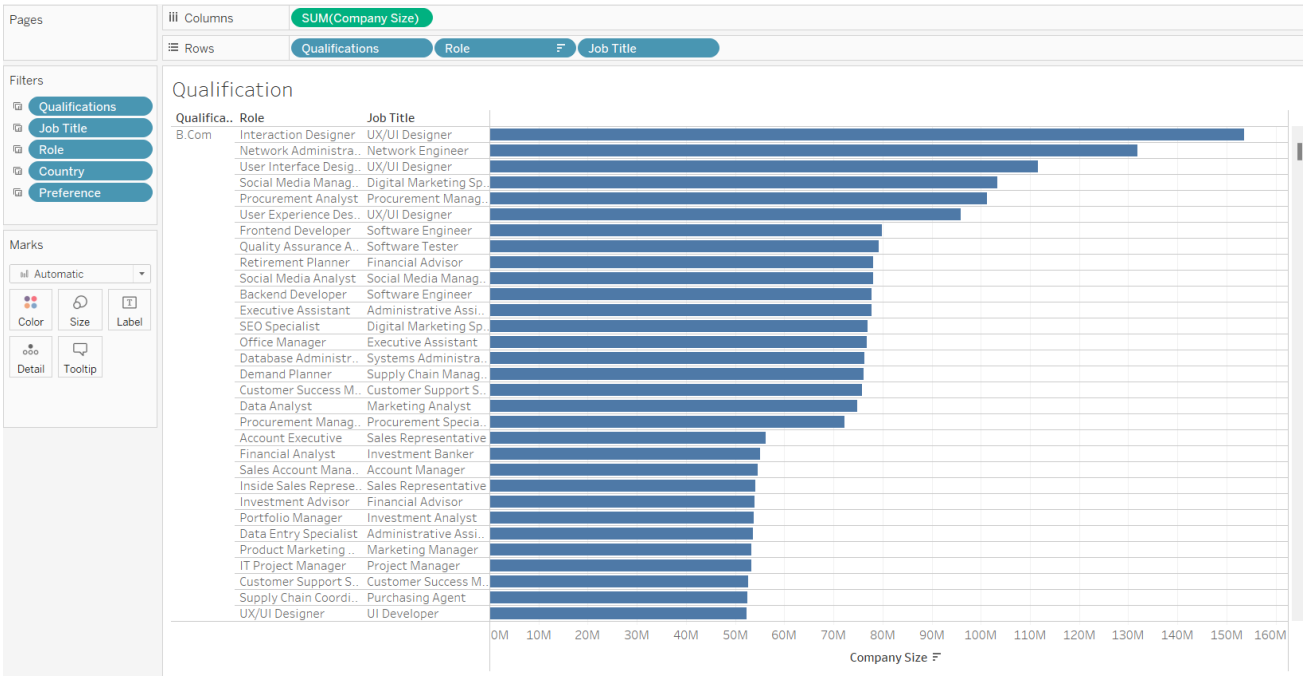
3. Training Sheet-3



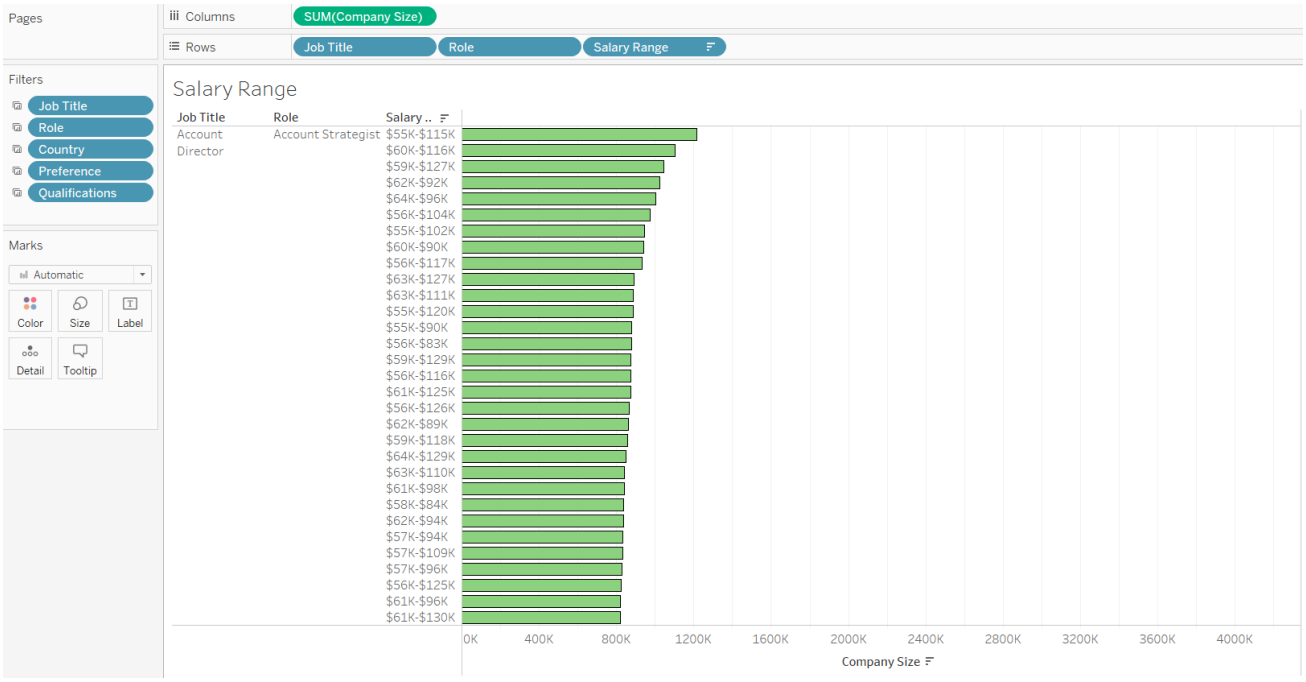
4. Training Sheet-4



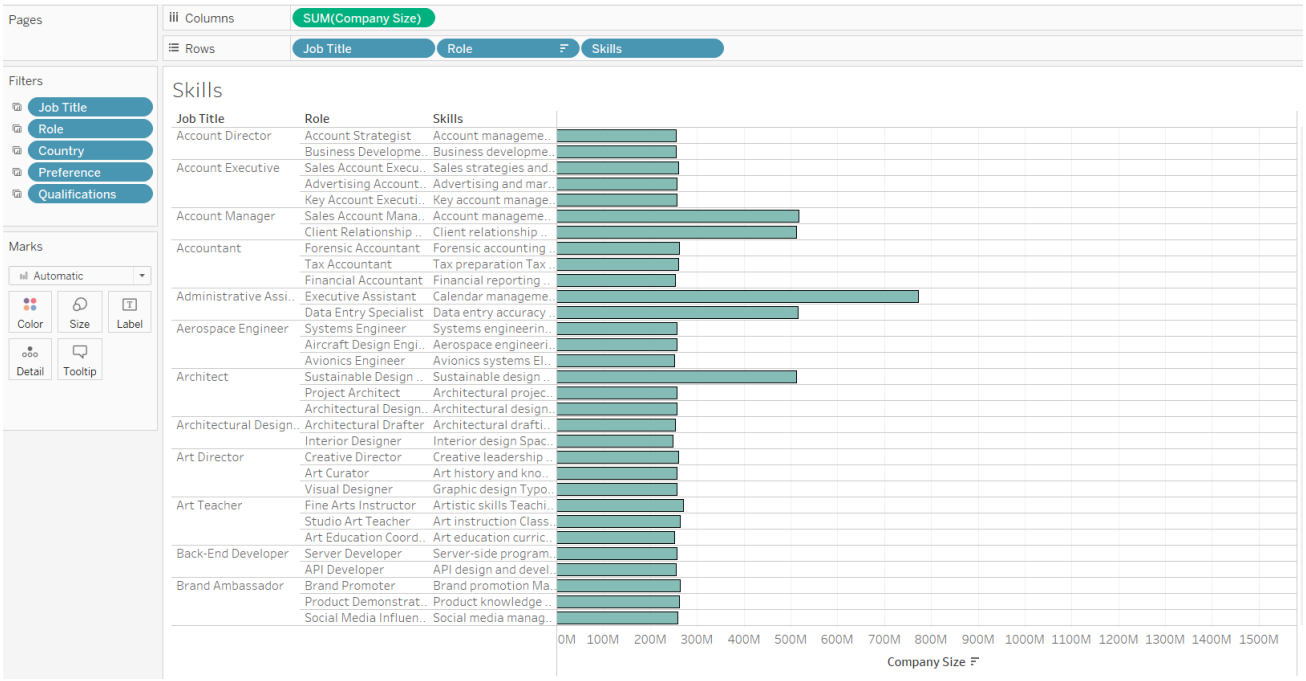
5. Training Sheet-5



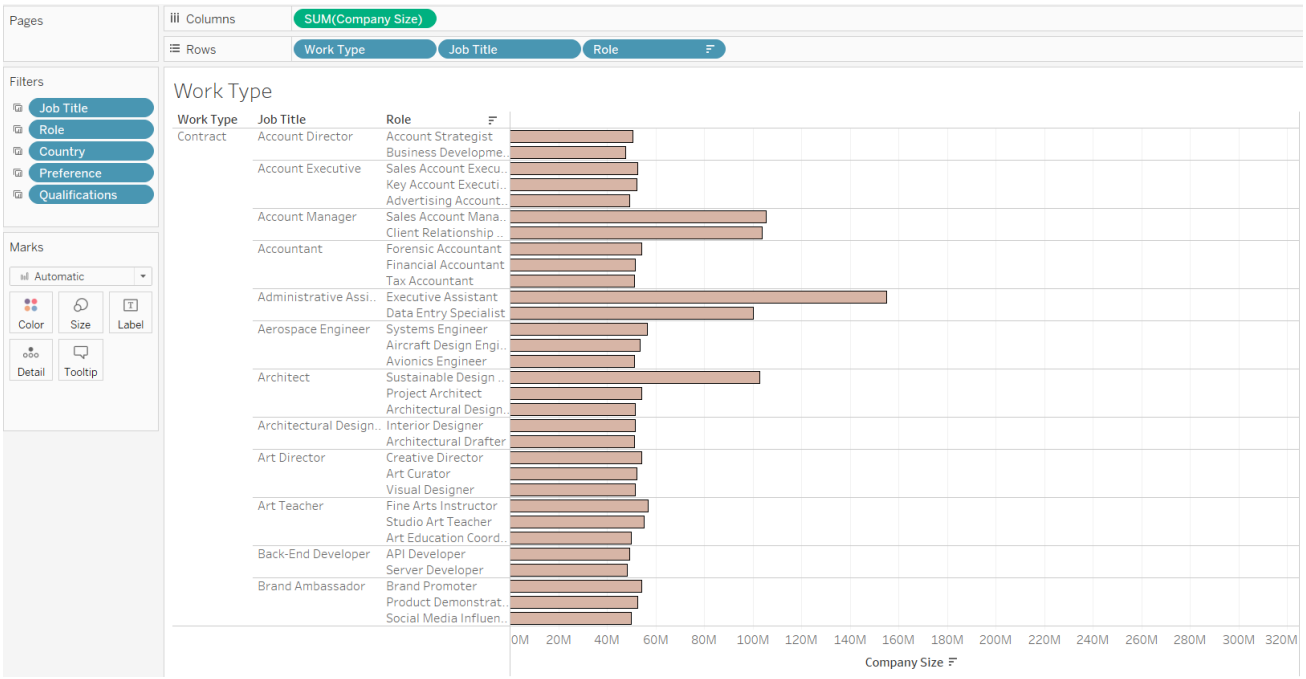
6. Training Sheet-6



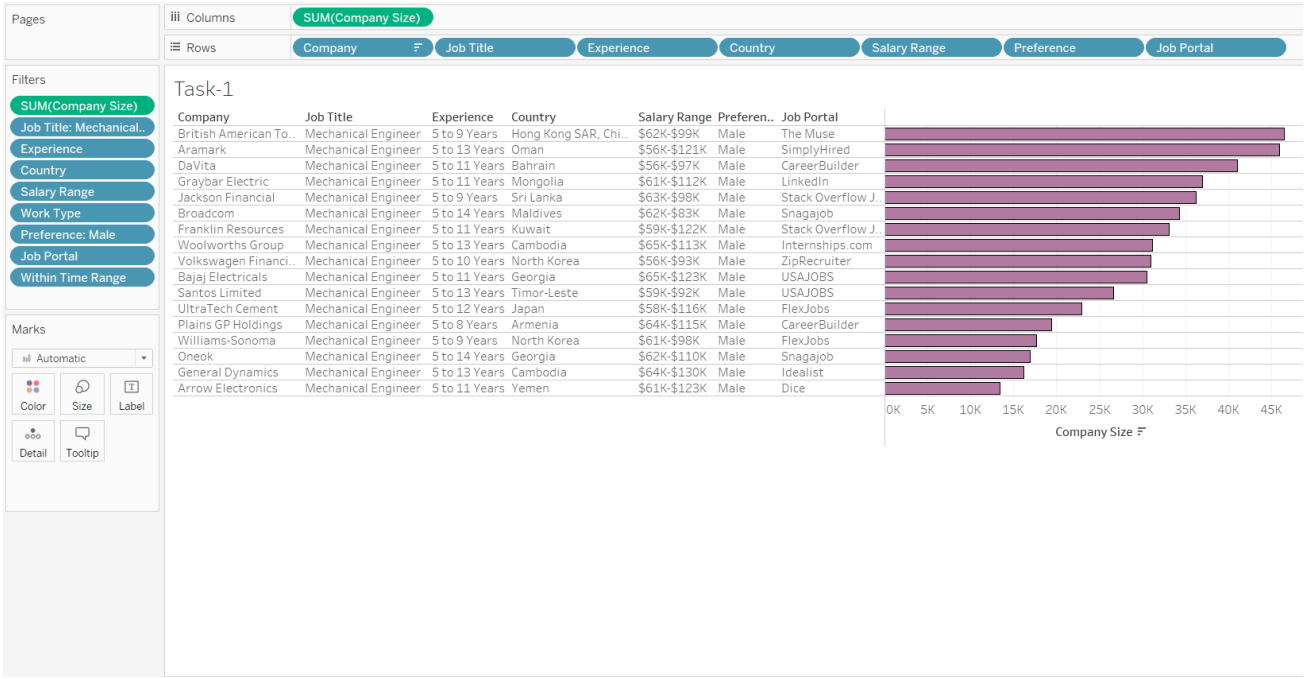
7. Training Sheet-7



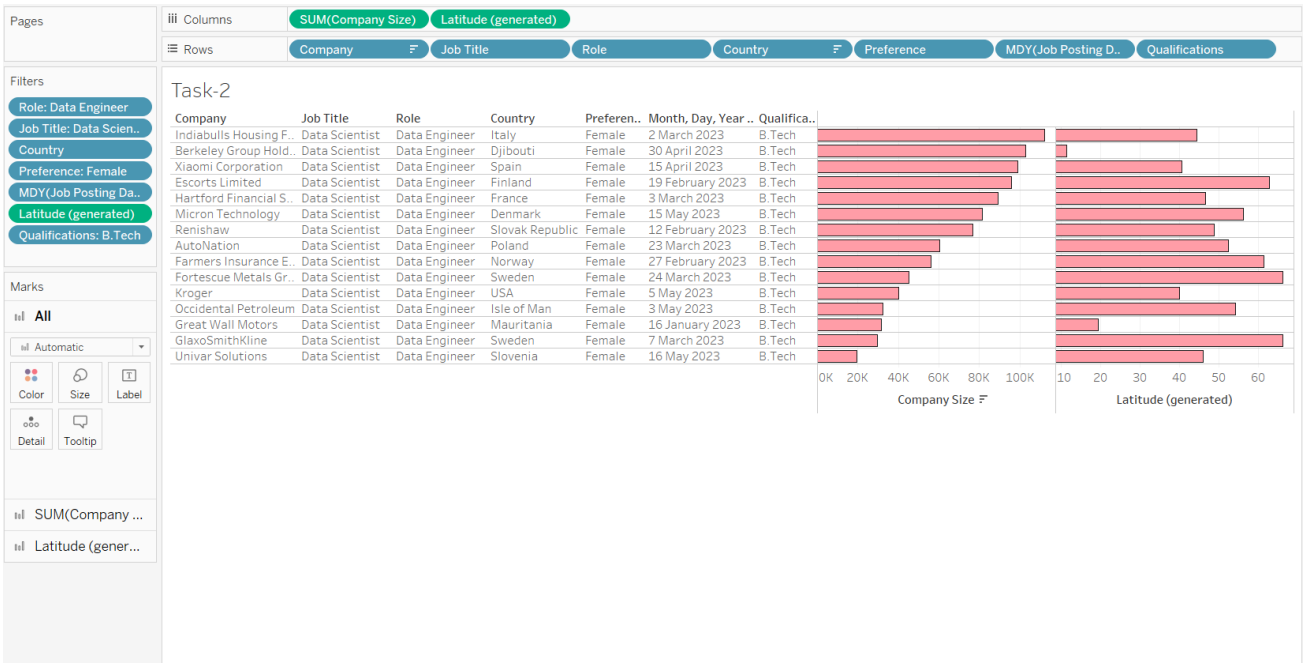
8. Training Sheet-8



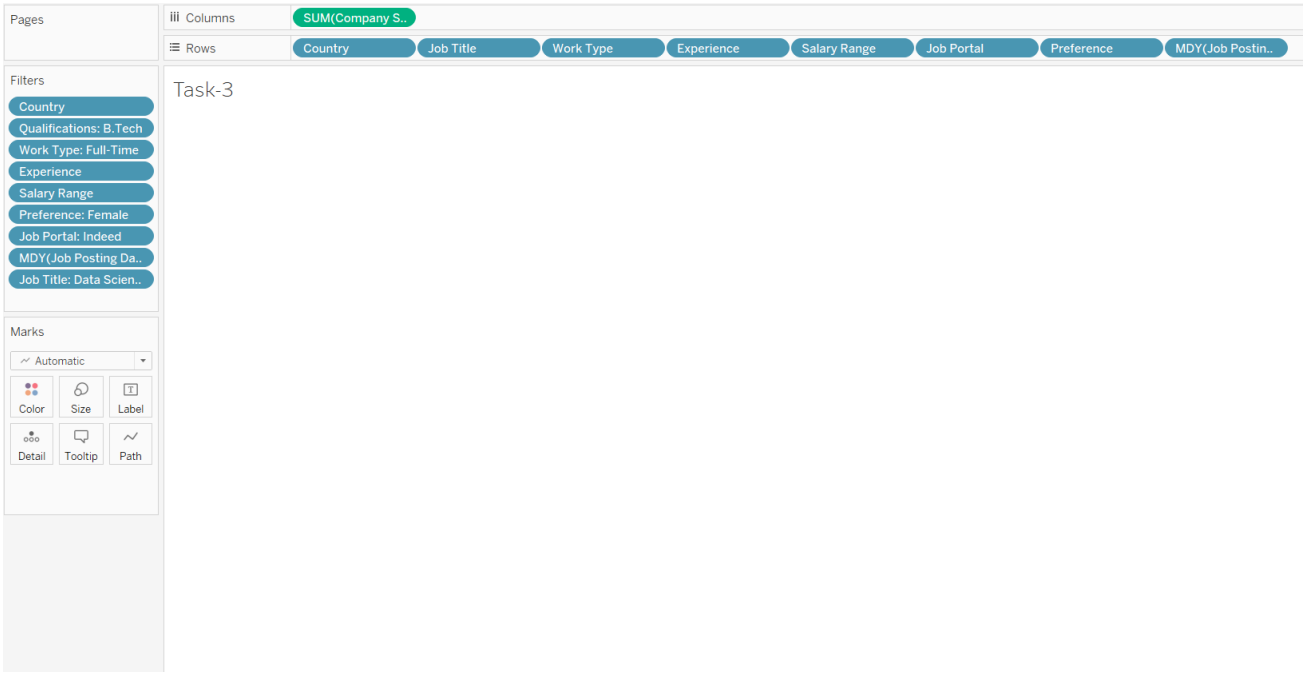
9. Internship Task-1



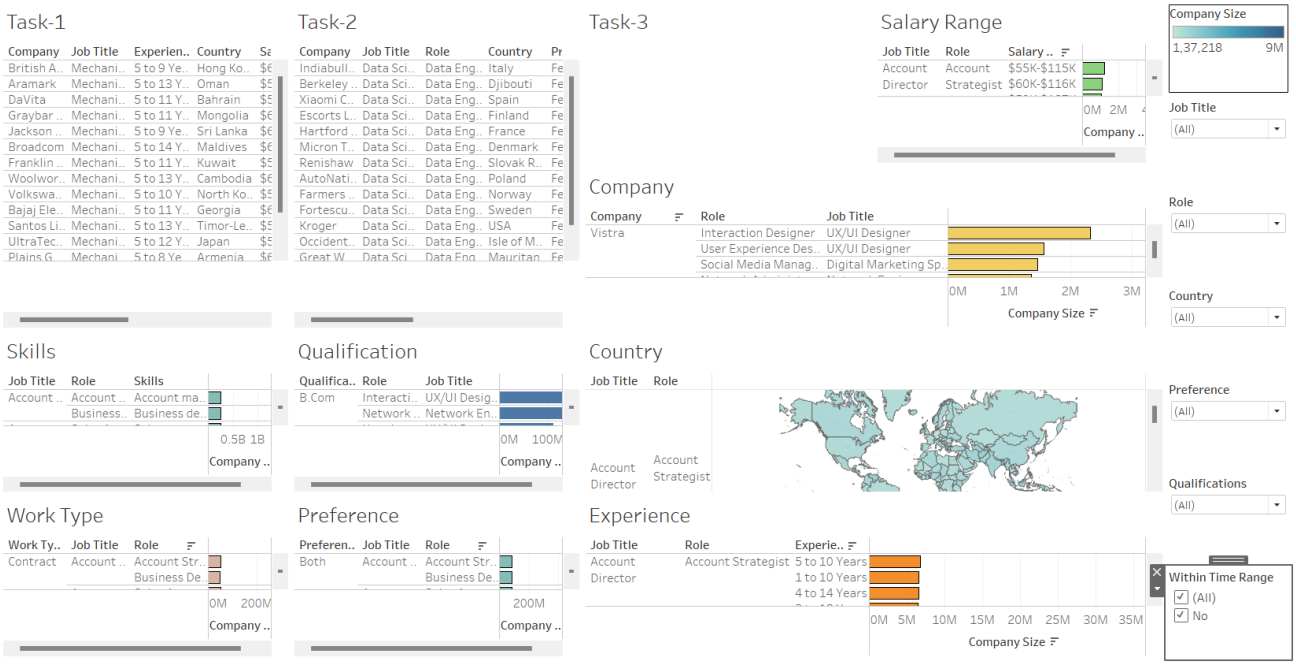
10. Internship Task-2



11. Internship Task-3



12. Internship Dashboard



Chapter 7 Challenges and Solutions

In the process of completing the tasks outlined in this report, several challenges were encountered due to the complexity of the requirements and the specificity of the data filters. Below is a detailed account of these challenges and the solutions implemented to overcome them.

Challenge 1: Complex Filtering Criteria for Task 1

Details: The first task required a chart to be drawn with very specific criteria, including filtering companies with less than 50,000 employees, job titles of Mechanical Engineer, over 5 years of experience, salaries over $50,000, and a work type that is either part-time or full-time. Additionally, candidates needed to have applied through a specific portal (Idealist), and the chart should only display data during a specific time frame (3 PM to 6 PM).

Solution: To address this, I utilized Python's Pandas library to implement the filters sequentially. I first filtered the dataset based on the core criteria (company size, job title, experience, etc.) and then added time-based filters using a custom function that only allows data to be processed within the specified time window. The final filtered dataset was used to create the required chart in Tableau. Time-dependent filtering was achieved by embedding a script that triggers chart updates within the specified hours.

Challenge 2: Exclusion of Specific Countries and Complex Geographic Filtering in Task 2

Details: Task 2 involved excluding Asian countries and countries starting with the letter ‘C,’ with additional latitude-based filtering. Furthermore, the role and job title filters added another layer of complexity.

Solution: I created a preprocessing script to exclude countries based on their names and geographic coordinates (latitude). The script was designed to run before data visualization to ensure only the relevant data was included. The use of regular expressions helped in filtering out countries starting with 'C,' and a combination of logical conditions was applied to handle the latitude constraints. This preprocessing was followed by the creation of a visualization in Tableau, which focused only on the required companies.

Challenge 3: Dynamic Colour Coding and Time-based Chart Updates in Task 3

Details: Task 3 required a chart to be created with dynamic colour coding based on the country (India and Germany) and required the colours to swap after 12 PM to 6 PM.

Solution: To achieve this, I used a conditional formatting feature within Tableau that dynamically changes the colour based on the time of day. I wrote a Python script to pre-process the data, ensuring

that it met all the criteria before passing it to Tableau. The time-based colour swap was implemented using a time-triggered update mechanism that refreshed the chart's visual properties at the designated time intervals.

Challenge 4: Integration of Multiple Data Sources and Tools

Details: The tasks required integration of various data sources and tools, including Kaggle datasets and Tableau for visualization, along with hosting the final dashboard on a website.

Solution: I ensured that all data cleaning and preparation steps were conducted in Python using Pandas and Numpy to ensure the data was ready for visualization. I then integrated the cleaned datasets into Tableau to create the required charts. Finally, I used Netlify to host the dashboard, ensuring that all dependencies were correctly linked and the dashboard was fully functional and responsive on the website.

Challenge 5: Ensuring Data Accuracy and Relevance

Details: Given the complex filters and criteria, there was a significant risk of either excluding relevant data or including irrelevant data, which could impact the accuracy of the charts.

Solution: To mitigate this, I implemented a step-by-step validation process at each stage of data filtering and processing. This involved checking intermediate outputs to ensure that each filter was applied correctly and the resulting data matched the specified criteria before moving on to the next step. Additionally, I cross-referenced the final output with the original dataset to ensure no critical data was omitted.

These challenges were significant, but through careful planning, rigorous testing, and the use of appropriate tools and techniques, I was able to successfully complete the tasks as required. The solutions implemented not only addressed the immediate challenges but also ensured that the final output was both accurate and aligned with the project objectives.

Chapter 8 Outcomes and Impact

The tasks outlined in this report have been completed successfully, demonstrating a comprehensive application of data visualization techniques and data filtering criteria. The primary outcomes include the creation of detailed charts that offer valuable insights into job trends based on specific parameters such as company size, job title, experience level, salary, and geographical location.

1. **Chart 1 - Mechanical Engineers in Asian Countries**:
    - **Outcome**: The chart was successfully created, showcasing the distribution of mechanical engineers across Asian companies with fewer than 50,000 employees. The data was filtered to meet specific criteria, including salary, work type, and candidate gender preference.
    - **Impact**: This visualization helps in identifying potential companies that align with the criteria, providing a focused approach for recruiters or analysts targeting this segment. The time-based filter ensures that the data remains relevant to specific time slots, aiding in time-sensitive decision-making.
2. **Chart 2 - Top 10 Companies for Data Engineers and Scientists**:
    - **Outcome**: A chart was generated highlighting the top 10 companies with the highest number of Data Engineers and Data Scientists, excluding Asian countries and certain geographic criteria. The preference for female candidates and specific qualifications were strictly adhered to.
    - **Impact**: This chart offers a strategic overview for talent acquisition teams focusing on non-Asian regions, enabling them to identify key companies that are likely to have a high demand for data-related roles. This can inform hiring strategies and workforce planning.
3. **Chart 3 - Data Science Jobs in India and Germany**:
    - **Outcome**: The chart successfully differentiated between Data Science jobs in India and Germany, with dynamic color-coding based on the time of day. The criteria for qualifications, experience, and salary were applied effectively.
    - **Impact**: This dynamic visualization provides an innovative way to compare job opportunities between two major markets, India, and Germany. The time-based colour swap feature enhances the usability of the chart, making it more interactive and relevant for users analyzing data during different times of the day.

Overall Impact: The completion of these tasks not only demonstrates technical proficiency in using Python libraries and Tableau for data visualization but also showcases the ability to interpret and apply

complex data filters to generate meaningful insights. The outcome, including the deployment of a web-based dashboard via Netlify, has broadened the accessibility of these insights, allowing stakeholders to interact with and utilize the data effectively in decision-making processes. This report and its outcomes are expected to contribute to more informed recruitment strategies and a deeper understanding of job market trends across different regions and job roles.

Chapter 9 Conclusion

In this report, I have detailed the task completion process for three complex data visualization tasks. Each task required precise filtering and charting based on specific criteria such as job titles, experience levels, geographic locations, and demographic preferences. These tasks involved the use of a job description dataset from Kaggle, with extensive data cleaning carried out using Python libraries like Pandas and NumPy.

The tasks were not only technically challenging but also required a deep understanding of data handling, filtering, and visualization. By employing these skills, I successfully created a Tableau dashboard that accurately represents the required data insights. Furthermore, I integrated these visualizations into a fully functional website, which was deployed using Netlify.

The report is structured into nine sections, covering every aspect from the initial background and objectives to the challenges encountered and the final outcomes. The successful completion of these tasks demonstrates a strong competency in data analysis, visualization, and web deployment, providing valuable insights into the data and highlighting my ability to manage and execute complex technical projects efficiently.

Link of the Website: https://ornate-kelpie-3a716c.netlify.app/