Data Analyst Job Postings

Adapa Aditya

CSC 457/557 (Data Analysis, Decision Making, and Visualization) – Project Topic Proposal – Fall 2023

Project Objective:

The project objective is to conduct a comprehensive analysis and visualization of a dataset comprising Data Analyst job postings, sourced from Kaggle. This dataset serves as a valuable resource, offering a wide range of information about job requirements, skills, demographic locations, and more, related to Data Analyst positions. The goal of this project is to extract actionable insights from this dataset, benefiting job seekers, hiring managers, and recruiters in the dynamic Data Analyst job market.

Data Source:

https://www.kaggle.com/datasets/lukebarousse/data-analyst-job-postings-google-search

Process of Data Analysis and Methods:

- **Data Gathering:** This project will show in detail the skills needed to become a Data Analyst, and various other details involving the Gsearch_jobs of Data Analyst Job Postings which was taken from Kaggle - considering various factors. This step involves collecting the dataset from Kaggle. This process includes data extraction and potential transformations to ensure that the data is suitable for analysis. Maintaining data integrity is a top priority during this phase.
- **Data Cleaning and Preprocessing:** Data quality is paramount for reliable analysis. This project will systematically address errors, inconsistencies, and data issues. Techniques such as data cleansing, filling in missing values, and standardization will be applied to ensure that the dataset is consistent and ready for exploration.
- **Data Exploration [EDA]:** The EDA phase will delve deep into the dataset. It includes activities like understanding the data's structure, identifying patterns, and performing in-depth analyses. Various statistical methods and visualization techniques will be employed to uncover insights that may not be immediately apparent.
- **Data Visualization:** Visual representations such as graphs, plots such as scatterplots, charts, and other plots will be used to provide a clear and intuitive view of the data. Additionally, using Microsoft Power BI to make the visualizations interactive.
- **Statistical Analytics:** Statistical analyses will be conducted to derive meaningful conclusions from the data.
- **Machine Learning Model:** As part of the project's advanced phase, a machine learning model will be trained using the dataset. Details regarding the type of model (e.g., classification, regression) and the specific algorithms chosen will be provided. The model aims to facilitate more informed decision-making and a deeper understanding of the dataset's complexities.

Project Purpose:

To derive key insights visually from Data Analyst Job Postings Dataset, such as, understanding job market trends, skills and qualifications, salary insights, demographic analysis, comparison between different job role names within Data Analysis, job posting sources, and predictive modeling using a machine learning model.

References:

https://github.com/pranavsai-98/Data-Analyst-Job-EDA/tree/master

https://github.com/MohamedMuneerM/naukri-data-analyst-job-posting-analysis#naukri-data-analyst-job-posting-analysis