

Exploratory Data Analysis of the Machine Learning & Data Science Industry

Title: Exploratory Data Analysis of the Machine Learning & Data Science Industry

Team:

- Param Chordiya
- Mihir Gujarathi

Introduction:

Our project aims to explore and decipher the rich dataset obtained from an extensive six-year industry survey focusing on data science and machine learning. This annual survey, conducted between 2017 and 2022, amassed a wealth of diverse responses, offering a global panorama of these rapidly evolving fields. Our objective is to delve deeper into this trove of information, uncovering captivating narratives and insightful trends within specific subsets of the data science community. Through a blend of narrative text and data exploration, our endeavor is to present compelling stories that reflect the experiences, preferences, and changes within various niches of the data science realm.

Problem Statement:

The project's primary focus revolves around understanding the evolving preferences and tool usage among diverse subsets of data science practitioners over the surveyed years. Specifically, we aim to investigate and track the changing landscape of tool adoption and preferences across different continents or regions. This exploration intends to provide a comprehensive understanding of how the utilization of tools in data science has evolved geographically, shedding light on trends and patterns unique to specific regions or communities.

Methodology:

- **Data Preprocessing:** We'll embark on a meticulous data cleaning and preprocessing phase to ensure the reliability and consistency of the survey responses.
- **Exploratory Data Analysis (EDA):** Utilizing EDA techniques, we'll unearth trends, distributions, and changes within subsets of interest, exploring shifts in tool usage and preferences.
- **Visualization:** Engaging visualization tools, we'll present our findings through graphical representations, facilitating a comprehensive understanding of the evolving dynamics.
- **Narrative Crafting:** The insights derived from our analysis will be woven into compelling narratives that highlight the unique journeys and transformations within the chosen subsets of the data science community.

Expected Outcome:

Our project aims to produce comprehensive narratives and visual representations that vividly illustrate the evolution of tool usage and preferences within diverse data science communities across different regions. By providing detailed insights and stories, we seek to contribute to a deeper understanding of the dynamic landscape of data science and machine learning over the surveyed years.

Conclusion:

In summary, our project endeavors to unravel and showcase the diverse stories embedded within the data science community. By combining rigorous data exploration with storytelling, we aim to present compelling narratives that shed light on the evolving trends, experiences, and preferences of various subsets within the data science realm.