Jordan Fuchs

Employee Satisfaction Analysis Report

**Executive Summary**

In this report, I present a comprehensive analysis of employee satisfaction within a fictional company using a dataset sourced from Kaggle. The overarching objective of this project was to unravel the intricate factors that shape employee satisfaction, uncover underlying patterns, and provide valuable insights to inform strategic decisions in the HR domain. To accomplish this, I harnessed the power of data science methodologies, proficiently leveraged visualization tools, and seamlessly integrated cloud technologies.

**1. Data Storage Choice: Data Warehouse**

The fundamental decision was to adopt a data warehouse approach for managing the dataset. This decision stemmed from the structured nature of the dataset, and Google Cloud Platform (GCP) emerged as the platform of choice due to its scalability, performance, and inherent compatibility with other GCP services. Specifically, Google BigQuery, a cloud-based data warehouse, emerged as the optimal solution for efficient data storage, retrieval, and analytical tasks. This choice was grounded in the recognition of the complexity of the dataset and the requirement for rapid and intricate querying - a characteristic often encountered in comprehensive employee satisfaction analyses.

**2. Integration with Cloud Services**

Seamless integration between the chosen data warehouse and the capabilities of Google Cloud Platform (GCP) was a pivotal step in the execution of this project. This integration enabled the streamlined processing and analysis of data by capitalizing on the distributed cloud infrastructure provided by GCP. This approach not only expedited data processing but also ensured data security and accessibility through GCP's robust cloud services.

**3. Spark Application Execution**

For effective data processing and analysis, I harnessed the capabilities of Apache Spark. To operationalize this, I authored a PySpark application that was tailored to the intricacies of our dataset. The PySpark application, a Python library for Spark, was engineered to seamlessly interface with Google BigQuery, retrieve the data, enact pertinent transformations, and synthesize meaningful insights regarding employee satisfaction.

**4. Documentation and Approach**

The strategic approach employed in this project can be encapsulated in the following key steps:

Data Collection and Exploration: The first step entailed procuring a comprehensive dataset from Kaggle. The dataset included crucial attributes such as age, gender, completed projects, productivity, satisfaction rates, feedback scores, department, position, joining date, and salary.

Data Preprocessing: To ensure the integrity of the analysis, the dataset underwent a meticulous preprocessing phase. This involved handling missing values, rectifying data types, and eliminating duplicate records.

Data Analysis: The cornerstone of the project was data analysis. I harnessed diverse data visualization techniques, ranging from scatter plots and histograms to correlation matrices. These tools enabled me to unearth latent relationships between employee attributes and their corresponding satisfaction rates.

Feature Engineering: Elevating the analysis, I conducted feature engineering by creating novel attributes, such as age groups and tenure, to amplify the scope of the investigation.

Hypothesis Testing: I subjected the dataset to statistical hypothesis testing to ascertain whether variables like gender, age, and department indeed wielded significant influence over satisfaction rates.

Visualization: To convey the findings effectively, I translated the analysis into insightful visualizations. Box plots, bar charts, and heatmaps were instrumental in illustrating intricate patterns and trends.

Spark Application: The core of the analysis lay in the PySpark application. This tailored script seamlessly interfaced with Google BigQuery, effectuated necessary transformations, facilitated comprehensive analyses, and culminated in the creation of enlightening visual representations.

**5. Results and Insights**

The in-depth analysis culminated in the following profound insights:

Productivity-Satisfaction: A notable positive correlation was discovered between higher productivity levels and enhanced satisfaction rates.

Gender and Department: The analysis highlighted discernible disparities in satisfaction rates influenced by factors like gender and department.

Tenure's Impact: Employees with prolonged tenures showcased markedly higher satisfaction levels, suggesting a correlation between longevity and contentment.

Salary and Satisfaction: A moderate positive correlation emerged between salary and satisfaction rates, indicating a noteworthy association.

**6. Conclusion**

This project served as a testament to the synergistic potential of data analytics and cloud technologies in unraveling the multifaceted realm of employee satisfaction dynamics. Through the confluence of a robust data warehouse, cloud services, and Spark applications, I seamlessly navigated data intricacies. The insights garnered from this analysis hold the potential to steer HR strategies towards augmenting employee satisfaction and holistic well-being.