**Notebooks as Business Reports - Enron Email Analysis**

Student’s Name

Institution

Course

Professor

Date

Notebooks as Business Reports - Enron Email Analysis

Introduction

The Enron Corporation, established in 1985, was an American energy, services, and commodities company. Although the company had increased market share, with a market value of over $70 billion, there was a dark reality of fraudulent accounting practices and financial scandals (Ali & El Gayar, 2019). The firm used deceptive activity to manipulate its financial statements and inflate its profits. The company's practices involved complex schemes and the creation of off-balance-sheet entities to hide the debt losses. The company collapsed in 2001 after the fraudulent activities were exposed. The Enron scandal also exposed weaknesses in corporate governance and raised concerns about the effectiveness of regulatory oversight. The Enron scandal is a cautionary tale about the importance of transparency, ethical behavior, and effective corporate governance in the business setting. The scandal highlighted the need for robust internal controls, independent audits, and a culture of integrity within organizations. Analyzing the Enron mail dataset, which consists of email communication from the firm's executives, employees, and associates, can provide insights into the company’s communication patterns and potentially uncover additional evidence related to scandals.

Methodology

Data Collection

The paper utilizes Enron as the major data source for the general paper analysis. It considers the firm's shortcomings which contributed to the company's collapse. Enron's database served this study with effective evaluation data for the general analysis. The data was also sourced from various text files used during the company's operation. For instance, data was extracted from JSON files, CSV, XML, and plain text files. Other business operation data was acquired from binary files like app data and media files like images, audio, video, and other compiled text.

*Example*

Import pyinputplus as pyip

Response = pip.inputNum()

Data Preparation

This process involves data cleaning to remove errors and duplicates when handling null values. Enron's collapse contributed to poor data preparation techniques, which led to increased financial scandals (Ali & El Gayar, 2019). In Python programming, pandas verify the dataset is complete and ready for further analysis. The panda's function is null () to complete data verification procedures.

Data Exploration

Data exploration gives a better understanding of the data for effective statistical and visual analysis to form a hypothesis and uncover potential patterns in the data. Also, panda, a Python library, is used for the exploratory portion of the email dataset. However, a mixture of *matplotlib, seaborn, and pandas* is used to create the visualization. The initial analysis and testing of the email dataset examination were developed in exploration\_development.ipynb, a notebook for explorations.

Analysis

Pandas allow the user to view the dataset, which comprises data presented in rows and columns with different datatypes like float, characters, and integers. Below is a sample of data analysis using Python.

import pandas as pd

columns = ['Sales,' 'Profits,' 'losses,' 'Expenses,' 'appreciation rate']

df = pd.read\_csv('Enron.csv', names=columns)

print ('Table 1: sales margin')

df["sales"].value\_counts()

Visualizations

In Enron's email dataset analysis, it is beneficial to use visualization techniques to visualize the data and the relationships between measurement and various datasets to see if certain patterns emerge. Data visualization is essential in making overall data analysis to observe various trends arising from various changes occurring due to various data changes witnessed. Enron's email dataset analysis used the *seaborn* pair plot, which pairs every feature with every other data feature, distinguished by an object type.

***Example***

import seaborn as sb

*# Pair Plot*

print('Figure 1: Pairwise plot of Enron Dataset')

sb.pairplot(df, ha='Species')

**Correlations**

This is the distinction between the sales and profits the company gains. The panda's correlation function tells more about this computation.

print('Table 4: Correlation Matrix of Enron Dataset')

df.corr()

The correlation can be visualized more easily using the heatmap below. The large red amount and the orange boxes show the high sales that are highly correlated with one another except for appreciation rates.

sb.heatmap(df.corr(), annot=True)

**Discussion**

The collected data samples show that the Enron’s collapse was greatly impacted by their email dataset corporation. All sales, profits, losses and purchase data should be recorded and well stored for future reference and decision-making purposes. The Enron’s databases were not well maintained since they did not clear duplicates and any other irrelevant data which could easily lead to the database breakdown. In an organization, database is a critical structure which need to be well-organized and maintained. The organization’s failure to monitor their email dataset, led to their immediate breakdown. The increase in the organization’s expenses with minimal profits contributed to their collapse. Also, database management need trustworthy leaders and employees who cannot delete or alter any information without legal authorization. The presence of corrupt and untrusted leaders contributed to the organization’s collapse.

**Conclusion**

The analysis covered in Enron's email dataset analysis uncovered many interesting details about Enron's datasets. The company's collapse was attributed to its failure to monitor its sales, promotion, profits, losses, and expenses data using data visualization, collection, preparation, and exploration techniques.

**Reference**

Ali, R. S. H., & El Gayar, N. (2019, December). Sentiment analysis using unlabeled email data. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 328-333). IEEE. <https://wwww.easychair.org/publications/preprint_download/HkRB>