Eulalia Rivera

Analytics Programming Tools

CIS 3330

Business Analytics Project

Summary

This is a summary of the business analytics (BA) report based on the company Vistra Corporation (2023) and formerly Vistra Energy Corporation. The BA process begins with collecting, sorting, and exploring large database organizations. BA consists of sequential applications that are descriptive, predictive, and prescriptive and are the major analytical components. Descriptive analytics analysis is based on graphic results including measures of central tendency (mean, median and mode) and measures of dispersion (standard deviation). Predictive analytics analysis is a more advanced method of data analysis that uses probabilities to make assessments of what could happen in the future. Predictive analysis uses that data and identifies correlations between variables. Knowing that such relationships exist explains why one set of independent variables influences a dependent variable. Accurate predictive analytics analysis helps business to reduce risks and costs.

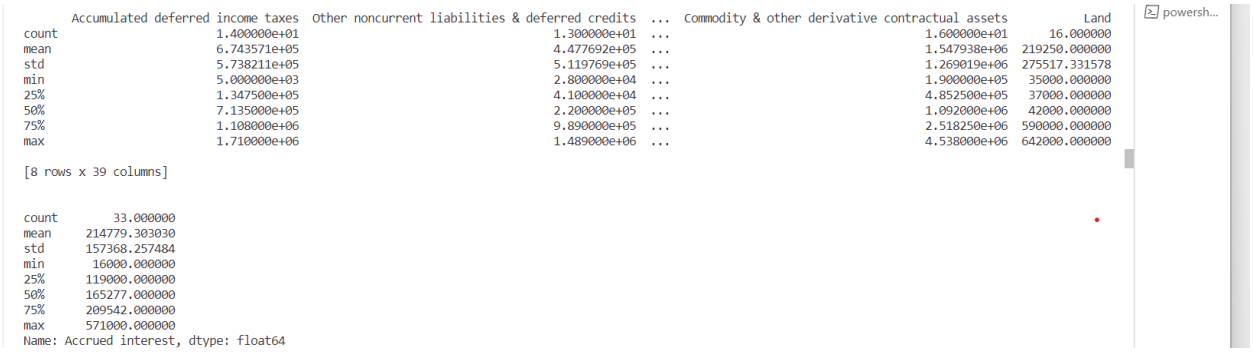
Descriptive Analytics

Descriptive Analytics is focused on describing the dataset. The main five steps of descriptive analytics process used in the dataset of Vistra Corp are: collection of data, data preparation, data cleaning, exploratory data, and data visualization. The data that I used was found in the UTEP database. Vistra Corp offers only data reports on the company main page. Collection of data, for this project I combined two sources of data, Vistra Corporation and the former company named Vistra Energy Corporation. After the aggregate operation, and for data

preparation I used a transpose function to switch the rows horizontally. For data cleaning, I removed errors, inconsistencies, and missing values so that it looked organized and ready to be analyzed. In explanatory data, I analyzed the data to find trends in patterns and relationships. Finally, data visualization is represented using graphs and charts to make it easy to understand.

Outputs

This output contains two results. The first result is for showing each column in the fata frame along measures of central tendency. The second output describes only one column of the data frame along measures of central tendency.



```
       Accumulated deferred income taxes  Other noncurrent liabilities & deferred credits  ...  Commodity & other derivative contractual assets          Land
count                       1.400000e+01                                    1.300000e+01  ...                                    1.600000e+01     16.000000
mean                        6.743571e+05                                    4.477692e+05  ...                                    1.547938e+06 219250.000000
std                         5.738211e+05                                    5.119769e+05  ...                                    1.269019e+06 275517.331578
min                         5.000000e+03                                    2.800000e+04  ...                                    1.900000e+05  35000.000000
25%                         1.347500e+05                                    4.100000e+04  ...                                    4.852500e+05  37000.000000
50%                         7.135000e+05                                    2.200000e+05  ...                                    1.092000e+06  42000.000000
75%                         1.108000e+06                                    9.890000e+05  ...                                    2.518250e+06 590000.000000
max                         1.710000e+06                                    1.489000e+06  ...                                    4.538000e+06 642000.000000

[8 rows x 39 columns]


count        33.000000
mean     214779.303030
std      157368.257484
min       16000.000000
25%      119000.000000
50%      165277.000000
75%      209542.000000
max      571000.000000
Name: Accrued interest, dtype: float64
```

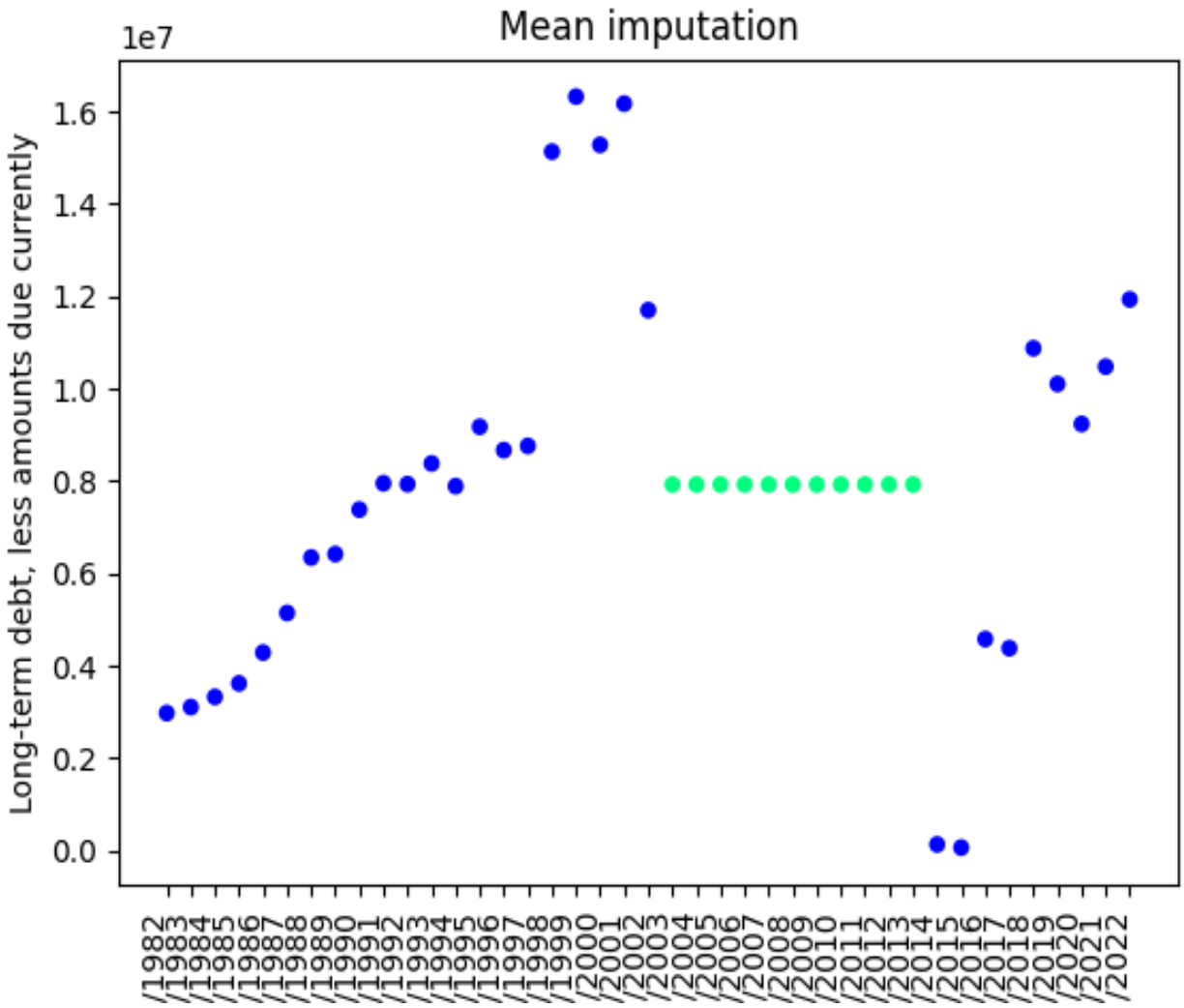
The third output is returning a new data frame where the data types have been changed to the float type. I coded the astype() method for this purpose.

```
Accrued interest                                          float64
Restricted cash.1                                         float64
Total liabilities                                         float64
Commodity & other derivative contractual liabilities.1    float64
Commodity & other derivative contractual liabilities      float64
Total current assets                                      float64
Common stock                                              float64
Short-term borrowings                                     float64
Materials & supplies                                      float64
Total current liabilities                                 float64
Uncertain tax positions, including accrued interest       float64
Long-term debt, less amounts due currently                float64
Commodity & other derivative contractual assets.1         float64
Total assets                                              float64
Accrued taxes                                             float64
Trade accounts payable                                    float64
Allowance for uncollectible accounts                      float64
Long-term debt due currently                              float64
Fuel stock                                                float64
Accumulated other comprehensive income (loss)             float64
Natural gas in storage                                    float64
Restricted cash                                           float64
Inventories                                               float64
Other noncurrent assets                                   float64
Less accumulated depreciation                             float64
Construction work in progress                             float64
Accumulated deferred income taxes.1                       float64
Electric                                                  float64
Investments                                               float64
Goodwill                                                  float64
Other current assets                                      float64
```

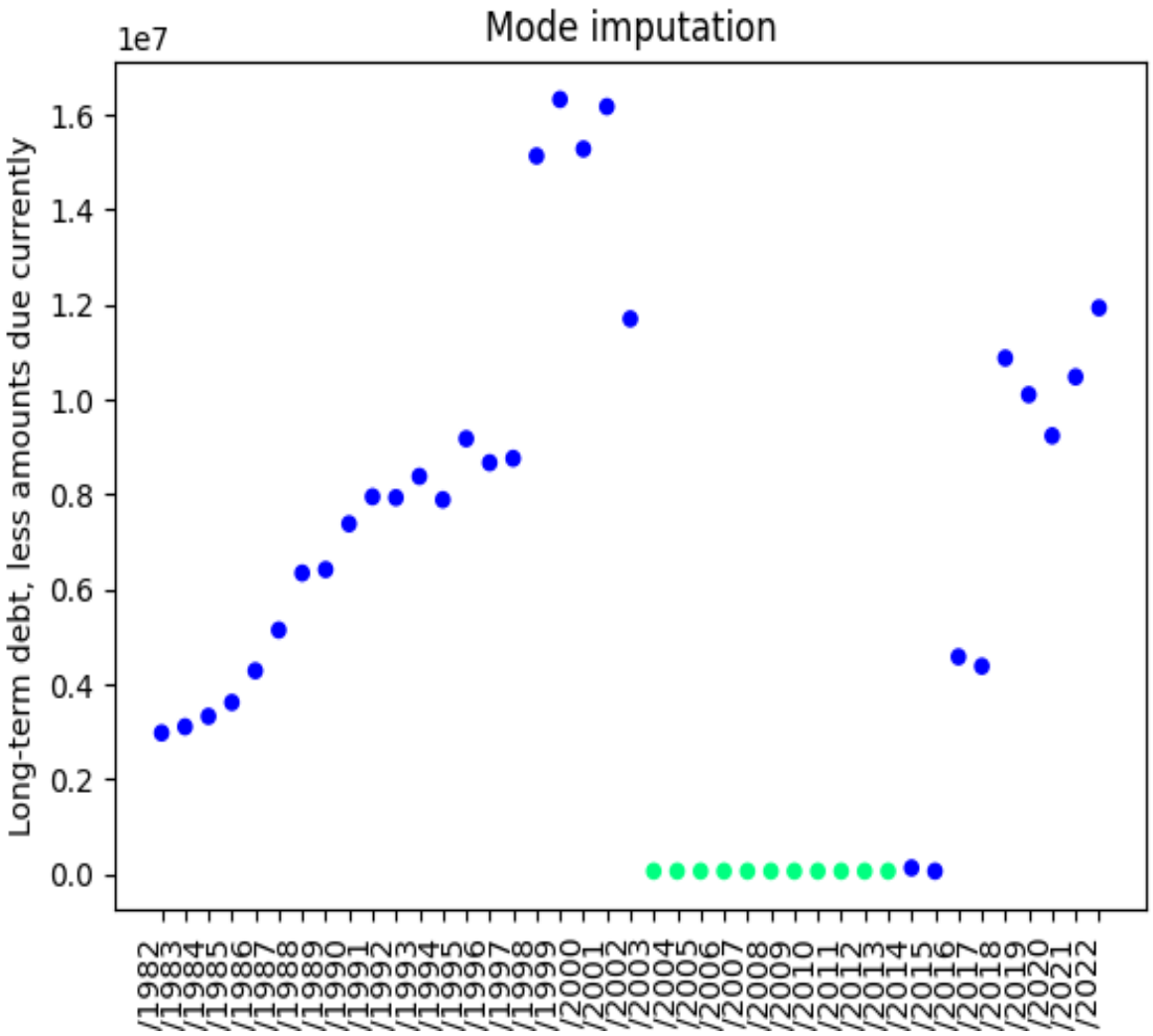
This output is counting the number of missing values in each column using the isna() function.

```
Commodity & other derivative contractual liabilities      25
Total current assets                                       0
Common stock                                               2
Short-term borrowings                                     32
Materials & supplies                                       0
Total current liabilities                                  0
Uncertain tax positions, including accrued interest       33
Long-term debt, less amounts due currently                11
Commodity & other derivative contractual assets.1         26
Total assets                                               0
Accrued taxes                                             34
Trade accounts payable                                     0
Allowance for uncollectible accounts                      26
Long-term debt due currently                               2
Fuel stock                                                 0
Accumulated other comprehensive income (loss)             21
Natural gas in storage                                    15
Restricted cash                                           21
Inventories                                               18
Other noncurrent assets                                   20
Less accumulated depreciation                              2
Construction work in progress                             11
Accumulated deferred income taxes.1                       36
Electric                                                  36
Investments                                               11
Goodwill                                                  15
Other current assets                                       6
Asset retirement obligations                              32
Trade accounts receivable - net                            0
Cash & cash equivalents                                    0
Other current liabilities                                  0
Commodity & other derivative contractual assets           25
Land                                                      25
dtype: int64
```

This graph is a Scatter Plot showing the mean imputation with the long-term debt and less amounts due currently replacing missing values with mean.



This graph is a Scatter Plot showing the mode imputation also with the long-term debt and less amounts due currently replacing missing values with mode.



Predictive Analytics

Predictive analytics is a more advanced method of data analytics as it makes predictions based on data. Predictive analytics looks like historical data and analyses past data trends to predict what will happen next. Like descriptive analytics, prescriptive analytics uses data mining and machine learning to predict opportunities in which the firm can take advantage. If predictive analytics is done correctly, predictive analysis can provide several benefits. Some benefits of choosing predictive analysis are the following: risk reduction, better strategy, and optimization. If Vistra Corporation chooses predictive analytics will create a more accurate forecast of demand for electricity and natural gas.

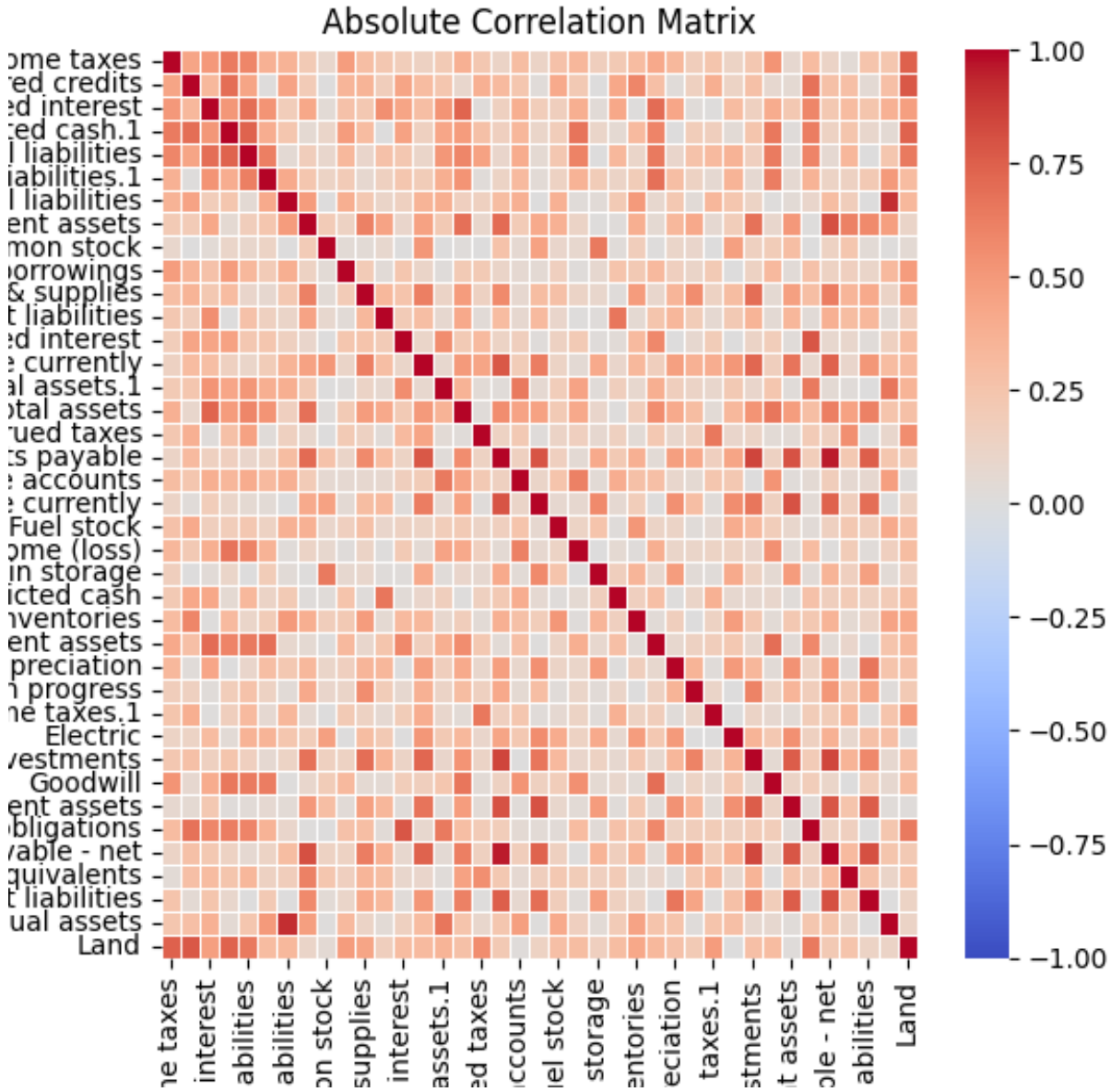
ANOVA (Analysis of Variance)

The ANOVA test whether any group is different from any other group, but it does not specify which group is different. So, printing a summary of the test just returns a single p value. The first step before to execute the ANOVA test is to define the null and alternate hypothesis. The null hypothesis test if is no significant difference among the groups, and alternate hypothesis test if is a significant difference among the groups. In general, if the p-value associated with the F is smaller than 0.05, then the null hypothesis is rejected and the alternative hypothesis is supported. If the null hypothesis is rejected, we can conclude that the means of all groups are not equal.

```
_____
                          sum_sq    df         F    PR(>F)
Q("Accrued interest")  4.618448e+13   1.0  8.146064  0.007626
Residual               1.757559e+14  31.0       NaN       NaN
PS C:\Users\erivera17\Documents\GitHub\final_project_calderon>
```

Heatmap of the Correlation of the Final_project_3 Data

The diagonal has elements with correlation 1 because every element is perfectly correlated with itself. Blue indicates highly negative correlation, red indicates highly positive correlation and white is not correlation. “Correlation does not mean causation.” In other words, just because two variables are correlated does not mean they influence each other.



Findings--- There is not findings.

In conclusion, descriptive and predictive analytics arranges as steps in the BA process can help Vistra Corp to find opportunities in data, predict trends that forecast future opportunities, and make more informed decisions. This process is challenging due to the amount of time that is dedicated to putting the data together and making it work properly. In the end, completing this research brought a lot of insight to analytics and the field of business.

References

Vistra Corporation. (2023, January 31). *Vistra*. Vistra Corp. https://www.vistracorp.com/