**Analysis Report**

**Title: Analysis of the Impact of Firm Characteristics on the Fog Index**

**Objective:**

The given project is designed to analyze the relationship between various firm characteristics (size, market\_to\_book, special\_items, complexity\_of\_operations, and incorporation\_state) and the fog index, a measure of financial disclosure readability. The analysis focuses on five companies (AAPL, AMZN, NFLX, GOOGL, and META) between 2012 and 2021.

**Data Preprocessing:**

The dataset is filtered to include only the data for the five specified firms (AAPL, AMZN, NFLX, GOOGL, and META) between 2012 and 2021.

New variables are calculated for the regression:

size: Log of total assets

market\_to\_book: Market value divided by stockholders' equity

special\_items: Operating income before depreciation minus net income (loss)

complexity\_of\_operations: Cost of goods sold divided by net sales/turnover

incorporation\_state: Assuming North American Industry Classification Code (NAICS) represents the incorporation state since it's not specified in the question.

The data is further filtered to include only the relevant columns for the regression analysis.

**Regression Model:**

An Ordinary Least Squares (OLS) regression model is used to analyze the relationship between the dependent variable (fog index) and the independent variables (size, market\_to\_book, special\_items, complexity\_of\_operations, and incorporation\_state).

The formula used for the regression is: fog ~ size + market\_to\_book + special\_items + complexity\_of\_operations + C(incorporation\_state)

**Results:**

R-squared: 0.630, indicating that the model explains 63% of the variance in the fog index.

Adjusted R-squared: 0.629, which is very close to the R-squared value, suggesting that the model has not been overfitted.

The F-statistic (456.5) and its associated probability (0.00) imply that the overall model is statistically significant.

The coefficients, standard errors, t-values, and p-values are provided for each of the independent variables, helping to interpret their individual relationships with the fog index.

Additional Details and Analysis of Regression Output for Each Variable:

Intercept (22.4093): This represents the baseline fog index value when all independent variables are zero. Given that some of the variables are logged or unlikely to be zero in practice, the intercept should not be interpreted literally.

Incorporation State (Categorical Variable):

C(incorporation\_state)[T.454110]: The coefficient of -0.7197 suggests that, on average, firms with NAICS code 454110 have a fog index lower by 0.7197 compared to the reference group (the group not included in the dummy variables), holding other factors constant. The p-value (<0.001) indicates that this relationship is statistically significant.

C(incorporation\_state)[T.518210]: Similarly, firms with NAICS code 518210 have a fog index lower by 2.0676 compared to the reference group, holding other factors constant. This relationship is also statistically significant (p-value < 0.001).

C(incorporation\_state)[T.519130]: Firms with NAICS code 519130 have a fog index lower by 1.2364 compared to the reference group, holding other factors constant. This relationship is statistically significant (p-value < 0.001).

C(incorporation\_state)[T.532282]: Firms with NAICS code 532282 have a fog index lower by 0.6092 compared to the reference group, holding other factors constant. This relationship is statistically significant (p-value < 0.001).

Size (4.614e-14): The coefficient suggests that a 1% increase in total assets (logged) is associated with an extremely small (practically zero) increase in the fog index, holding other factors constant. However, the p-value (1.000) indicates that this relationship is not statistically significant.

Market\_to\_book (3.282e-15): The coefficient implies that a one-unit increase in the market-to-book ratio is associated with an extremely small (practically zero) increase in the fog index, holding other factors constant. The p-value (1.000) indicates that this relationship is not statistically significant.

Special\_items (7.431e-17): The coefficient suggests that a one-unit increase in special items is associated with an extremely small (practically zero) increase in the fog index, holding other factors constant. The p-value (1.000) indicates that this relationship is not statistically significant.

Complexity\_of\_operations (8.149e-14): The coefficient implies that a one-unit increase in the complexity of operations is associated with an extremely small (practically zero) increase in the fog index, holding other factors constant. The p-value (1.000) indicates that this relationship is not statistically significant.

In summary, the regression output indicates that the incorporation\_state variable has a statistically significant impact on the fog index. However, the other continuous variables (size, market\_to\_book, special\_items, and complexity\_of\_operations) do not show a statistically significant relationship with the fog index. The extremely small coefficients and high p-values for these variables suggest that they may not be meaningful predictors of the fog index in the given model.

**Caveats and Limitations:**

The condition number is large (1.59e+05), suggesting that there might be strong multicollinearity or other numerical problems in the dataset. This could impact the reliability of the regression coefficients and should be investigated further.

It is essential to note that the analysis assumes that the NAICS code represents the incorporation state, which might not be accurate. If the actual incorporation state is available, it should be used instead.

**Conclusion:**

The OLS regression model provides insight into the relationship between the firm characteristics and the fog index. The model has a relatively high R-squared value, indicating that it can explain a substantial portion of the variation in the fog index. However, the large condition number suggests that multicollinearity or other numerical issues might be present in the data, which should be further investigated.