Freelancer Earnings Model Report (using IBM SPSS Modeler)

By – Arijit Bhadra | MBA (Business Analytics)
Link to: GitHub, LinkedIn

About: In this project, I leveraged IBM SPSS Modeler to build a predictive analytics model aimed at forecasting Freelancer Earnings (USD) based on multiple inputs — a real-world scenario relevant in the gig economy and digital marketplaces.
Business Objective: To predict freelancer earnings based on inputs like job performance, platform, region, client rating, hourly rate, marketing spend, and more.

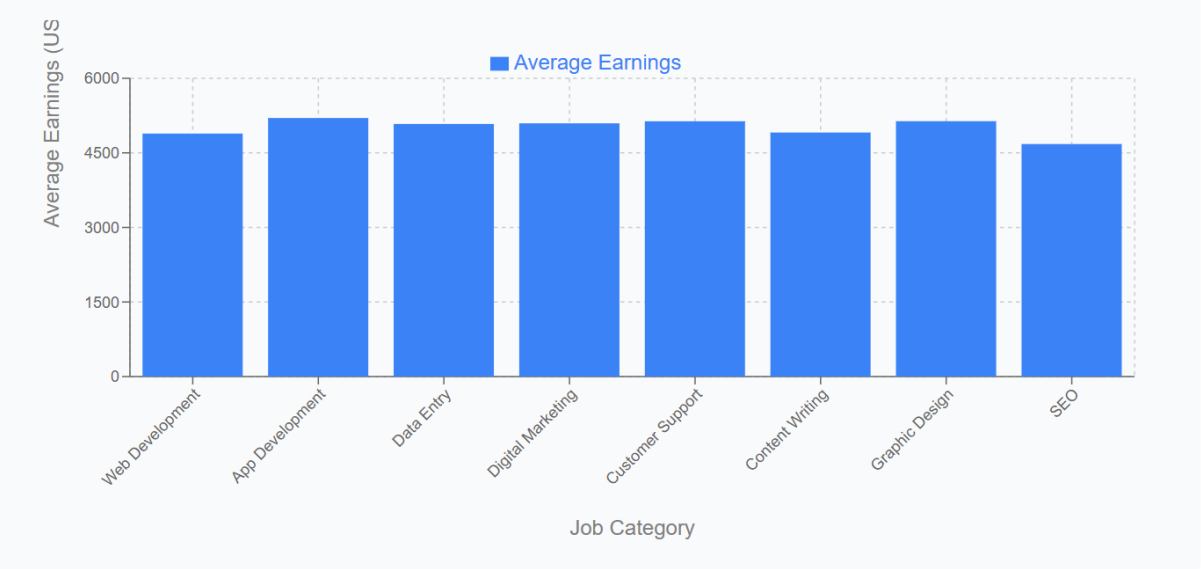
Dataset Highlights:

- **Records:** 1950 freelancers
- **Features:** 16 columns including categorical (Job Category, Platform), numerical (Hourly Rate, Client Rating), and derived metrics
- **New Feature Added:** Marketing Efficiency (Earnings / Marketing Spend)

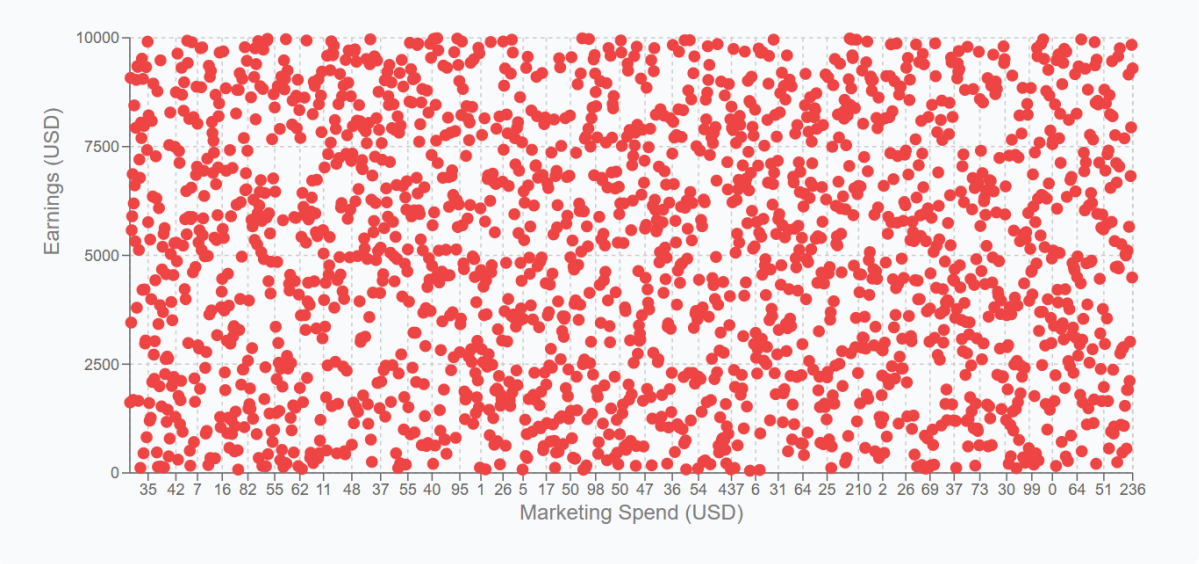
Data Preparation:

- Used **Auto Data Prep**, **Filter**, and **Derive** nodes in SPSS Modeler
- Cleaned, transformed, and engineered variables to improve model performance

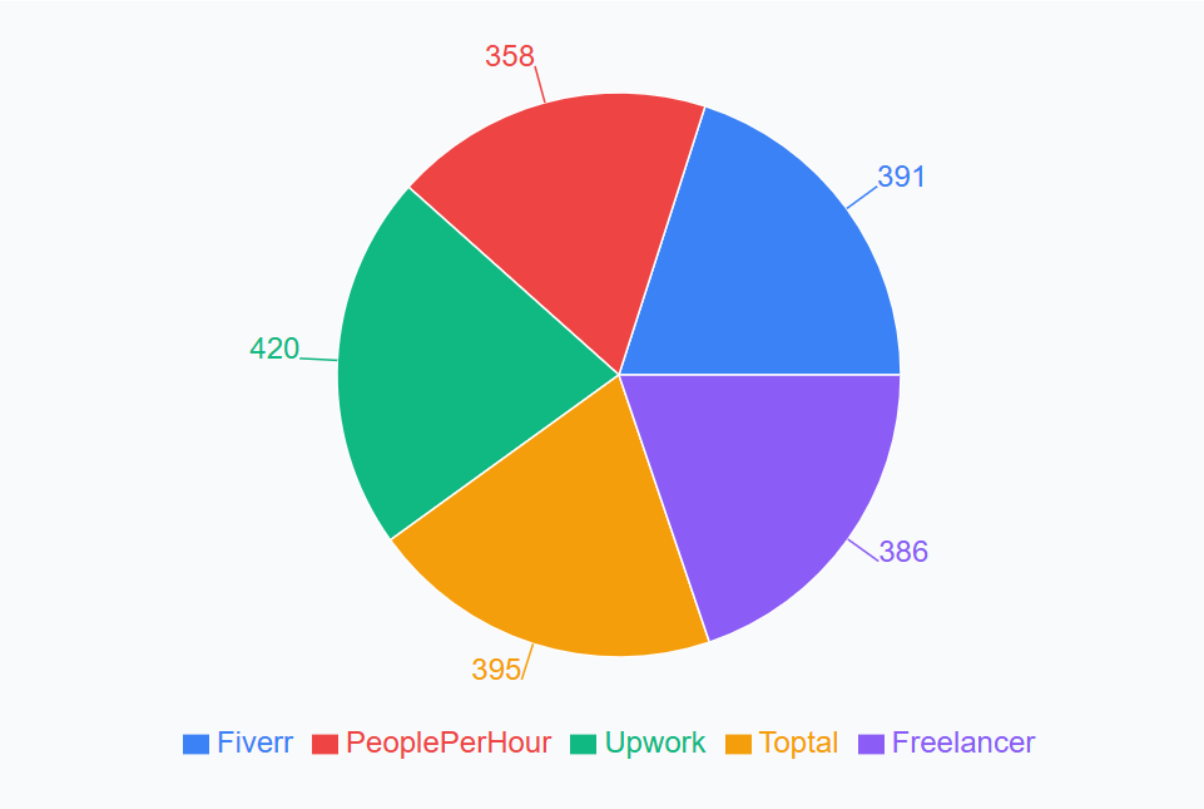
Data Visualisations:
1. Average Earnings by Job Category:



2. Earnings vs Marketing Spend:



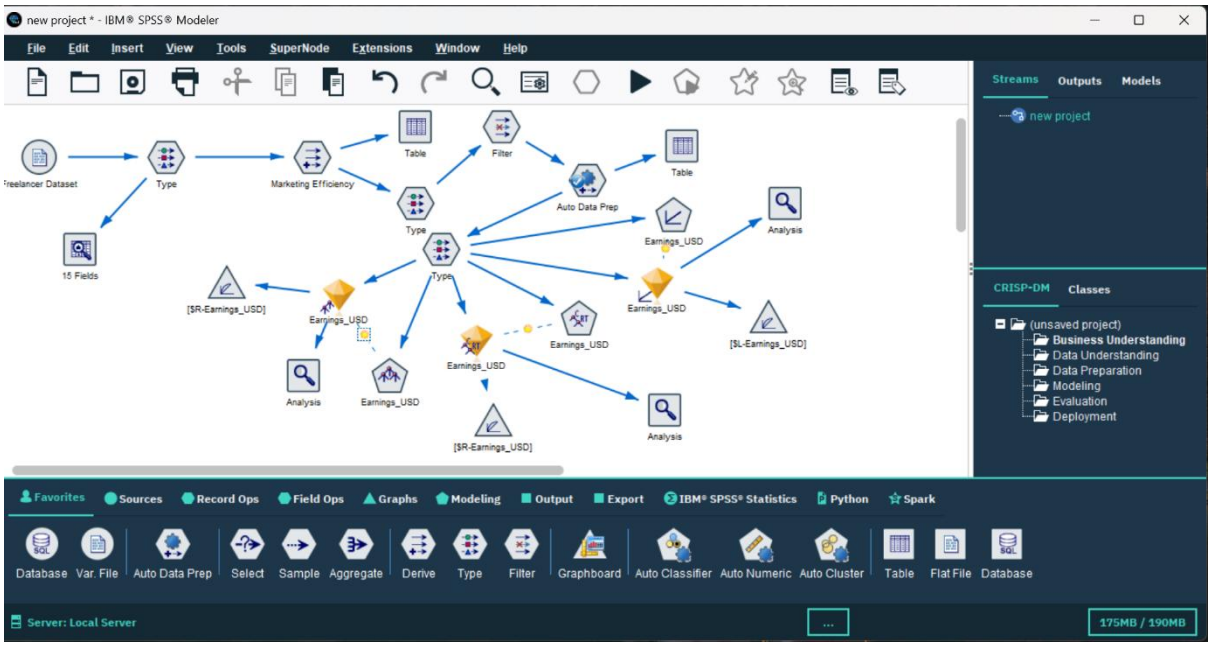
3. Distribution of Jobs by Platform:



IBM SPSS Modeler: This project involved building a predictive model in IBM SPSS Modeler to forecast freelancer earnings. The workflow included comprehensive data cleaning, use of Auto Data Preparation, variable filtering, and the creation of a derived

feature — Marketing Efficiency (Earnings ÷ Marketing Spend). Multiple models were developed, including C&R Tree, Tree-AS, and Linear Regression, with performance evaluated using Linear Correlation. The C&R Tree model showed the highest predictive accuracy (R = 0.932) and was selected for final analysis. The project demonstrates practical use of SPSS Modeler for forecasting in real-world business scenarios.

Stream (workflow – IBM SPSS Modeler):



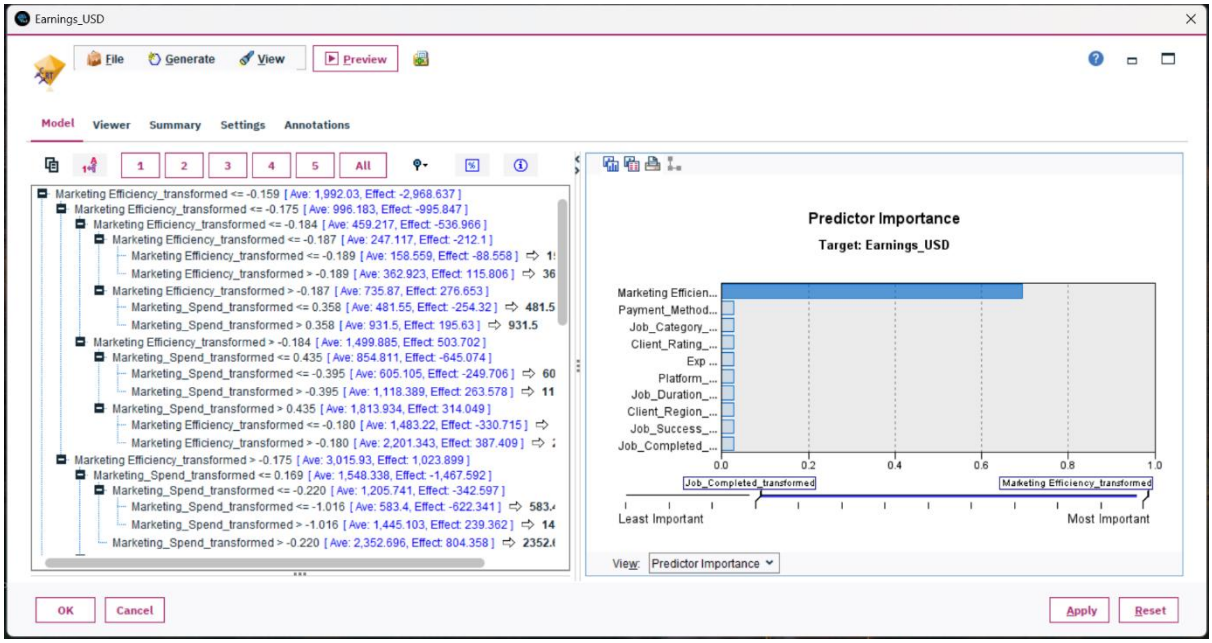
Model Comparison:

Model	Linear Correlation
C&R Tree	0.932
Tree-AS	0.897
Linear Regression	0.494

The performance of the three predictive models — C&R Tree, Tree-AS, and Linear Regression — was evaluated using the **R^2** (**coefficient of determination**) metric. This metric represents the proportion of variance in the target variable (freelancer earnings) that the model is able to explain based on the input features. A higher R^2 value indicates better predictive power and a stronger fit to the data.

Among the models, the C&R Tree performed the best, achieving an R^2 of 0.869, meaning it could explain approximately 86.9% of the variance in earnings. The Tree-AS model followed closely with an R^2 of 0.805 (80.5%), while Linear Regression showed significantly lower predictive ability, with an R^2 of just 0.244 (24.4%). Based on these results, the C&R Tree was selected as the final model due to its superior accuracy and interpretability.

C&R Tree: The C&R Tree model (Classification and Regression Tree) provided a detailed decision structure for predicting freelancer earnings. As shown in the model output, Marketing Efficiency emerged as the most significant predictor, followed by factors like Payment Method, Job Category, and Client Rating. The tree branches clearly illustrate how different ranges of these variables influence earnings. The Predictor Importance chart further confirms the dominant role of marketing-related variables in determining income, offering actionable insights for freelancers looking to optimize their performance and earnings strategy.



Key Insight: The analysis revealed that variables such as hourly rate, number of jobs completed, client rating, and marketing efficiency (earnings relative to marketing spend) were strong predictors of freelancer earnings. Additionally, certain platforms and regions consistently showed higher income potential, highlighting the influence of market dynamics and platform-specific demand.

Conclusion: This project demonstrates the effectiveness of decision tree-based models like **C&R Tree** in predicting freelancer earnings with high accuracy. By combining automated data preparation, feature engineering, and model comparison in IBM SPSS Modeler, we were able to develop a robust forecasting model that provides actionable insights for freelancers and platforms alike. This approach can be extended to other gig economy datasets for strategic decision-making and income optimization.

GitHub Repository: Click here