

# Case Study: Model Engineering

**Task**

**Credit Card Routing for Online Purchase via Predictive Modelling**

**Author: Supervised by:**

**Haribabu Vinoth. Sahar Qaadan**

Univ Year 2022-2023

Date: Mon, 25th Sep 2023

**Abstract**

Efficient online payment processing is of utmost importance in the dynamic realm of online retail, as it plays a crucial role in attaining consumer happiness and fostering company expansion. Nevertheless, several retail enterprises encounter the predicament of heightened rates of failure in online credit card transactions, resulting in significant financial consequences and negatively impacting consumer satisfaction. This study explores the development of a novel solution aimed at addressing these challenges: a predictive model designed to intelligently route credit card transactions in order to optimize Payment Service Providers (PSPs). Through the use of previous transaction data and the implementation of sophisticated machine learning methodologies, we have successfully devised a model with the ability to ascertain the optimal payment service provider (PSP) for every individual transaction. This technique not only offers the potential for a substantial increase in the success rate of online payments but also promotes a more streamlined and financially advantageous retail operation. The project is based on the well-regarded CRISP-DM methodology and encompasses many stages including business knowledge, data handling, modeling, and deployment. Its ultimate goal is to bring about a significant change in the field of digital transactions.

**Keywords:** Online Retail, Payment Processing, Payment Servise Providers, Machine Learning, Historical Transaction Data, Crisp-DM Methodology, Logistic Regression, Random Forest Model, Success Rate, Accuracy

Table of figures

[Figure 1: Provided Dataset 10](file:////Users/mac/Downloads/Feuille_de_style_PFA_Ing_INPT_18.06.21-6%20(1).docx#_Toc146465404)

[Figure 2: Data Quality Assessment 11](file:////Users/mac/Downloads/Feuille_de_style_PFA_Ing_INPT_18.06.21-6%20(1).docx#_Toc146465405)

[Figure 3: Data preparation 12](#_Toc146465406)

[Figure 4: Data Split 12](#_Toc146465407)

[Figure 5: Distribution of Transaction Amounts 14](#_Toc146465408)

[Figure 6: Success Rate by Country 15](#_Toc146465409)

[Figure 7: Success Rate Base on Card Type 16](#_Toc146465410)

[Figure 8: PSP Usage Distribution 17](#_Toc146465411)

[Figure 9: Impact of 3D security on Transaction Success 18](#_Toc146465412)

[Figure 10: Prepared Dataset 20](#_Toc146465413)

[Figure 11: Logistic Regression Model 21](#_Toc146465414)

[Figure 12: Random Forest Model 23](#_Toc146465415)

[Figure 13: Hyperparameter Optimization 24](#_Toc146465416)

Table des matières

[Case Study: Model Engineering 1](#_Toc146465980)

[Table of figures 3](#_Toc146465981)

[Table des matières 4](#_Toc146465982)

[Introduction: Project Planning 6](#_Toc146465983)

[**Chapter 1** 10](#_Toc146465984)

[1.1 Dataset Description: 11](#_Toc146465985)

[1.2 Data Quality Assessment: 12](#_Toc146465986)

[1.3 Data Preparation Techniques: 12](#_Toc146465987)

[1.4 Data Split: 13](#_Toc146465988)

[**Chapter 2** 14](#_Toc146465989)

[**Initial Data Analyses** 14](#_Toc146465990)

[Initial Data Analyses 15](#_Toc146465991)

[2.1 Distribution of Transaction Amounts 15](#_Toc146465992)

[2.2 Success Rates by Country 16](#_Toc146465993)

[2.3 Success Rate Based on Card Type 17](#_Toc146465994)

[2.4 PSP Usage Distribution 17](#_Toc146465995)

[2.5 Impact of 3D Security on Success 18](#_Toc146465996)

[**Chapter 3** 20](#_Toc146465997)

[3.1  Data Preprocessing 21](#_Toc146465998)

[3.2  Logistic Regression Model 22](#_Toc146465999)

[3.1  Accurate Predictive Model: Random Forest 23](#_Toc146466000)

[3.2  Implementation of the Model 24](#_Toc146466001)

[3.3  Improving the Random Forest Model 25](#_Toc146466002)

[**Chapter 4** 27](#_Toc146466003)

[**Operationalizing the Transaction Success Model for Everyday Business Functions** 27](#_Toc146466004)

[4.1  Seamless Deployment & Fluid Integration: 28](#_Toc146466005)

[4.2  Dynamic Recommendation Mechanism: 28](#_Toc146466006)

[4.3  Immersive GUI Experience: 28](#_Toc146466007)

[4.4  Proactive Monitoring & Analytics Suite: 29](#_Toc146466008)

[4.5  Iterative Model Refinement: 29](#_Toc146466009)

[4.6  Focused Error Scrutiny & Evolution: 29](#_Toc146466010)

[Conclusion 30](#_Toc146466011)

[Bibliography 32](#_Toc146466012)

[**Annexes** 33](#_Toc146466013)

[**Annexe A : Full Code** 33](#_Toc146466014)

# Introduction: Project Planning

In a time characterized by the widespread availability of data, our ability to extract valuable information, make informed choices, and automate tasks using machine learning techniques is unmatched. As several businesses, ranging from healthcare to e-commerce, use the capabilities of machine learning, they underscore its inherent worth in facilitating optimization and cultivating creativity. Primarily, the exploration of machine learning is driven by a solitary aim: to uncover underlying patterns inside data, using them to advance intelligent predictions and automations.

The complex procedure involved in initiating a machine learning project has resemblance to the artistic process of molding raw stone. The process starts with the careful compilation and refinement of data, establishing the fundamental basis for the whole undertaking. The subsequent step involves a meticulous process of identifying and modifying features, with the aim of equipping the algorithms with the most significant predictors. Utilizing an extensive collection of patterns as a basis, models undergo a rigorous training process, whereby they adapt and refine their comprehension until they achieve or exceed certain criteria. However, the process of creating a model is not the last stage of its journey. The model's effectiveness is ultimately evaluated in real-world scenarios, where it must make critical real-time deductions that might have substantial implications for business outcomes.

In a technologically advanced setting that offers programming tools such as Python and R, the process of creating machine learning models becomes less arduous and more of a journey of discovery, facilitated by dedicated libraries and frameworks. Nevertheless, it is crucial to bear in mind that the foundation of these activities relies heavily on a resilient data infrastructure, which guarantees the smooth flow of extensive information and complex calculations.

In the realm of online shopping, where each click has the potential to result in a transaction, it is of utmost importance to prioritize the establishment of seamless online payment processing. The emergence of difficulties such as increasing instances of credit card payment failures presents a situation that undermines both client confidence and financial performance. This project navigates over the aforementioned problems by using predictive modeling as its guiding tool. What is our mission? The objective is to transform the credit card routing mechanism from a rigid, manual procedure governed by predefined rules to a sophisticated, adaptable system powered by data analysis. By using a collection of past transaction narratives, our objective is to enhance the system's ability to identify the most suitable Payment Service Provider (PSP) for each transaction. The goal is clear and has significant implications: to increase the rates of successful payments while carefully avoiding expenses.

As participants in this expedition, working with the pioneers of the online payment industry, our goal is to develop a predictive framework that considers many factors, such as transactional timestamps and historical performance indicators of payment service providers (PSPs) like Moneycard and Simplecard. The primary objective is to cultivate a transactional setting in which errors are infrequent rather than typical. Each conducted transaction should not only be seen as a routine procedure, but also as a strategic choice based on data analysis. This approach aims to enhance client experiences and strengthen our position in the digital marketplace.

**Exploring the CRISP-DM Framework: An In-depth Analysis**

In commencing our endeavor, we use the esteemed **CRISP-DM (Cross-Industry Standard Process for Data Mining) framework**, renowned for its thorough, methodical, and iterative approach specifically designed for data-driven initiatives. The cyclical structure of this framework guarantees that when progress is made, each phase contributes to the subsequent one and, when needed, offers valuable insights to reassess and improve prior processes for the most favorable results.

company Understanding: Initially, our primary focus is to get a thorough understanding of our company goals and the obstacles we face. Through active involvement with stakeholders, namely those from the online payment department, we carefully analyze and extract their challenges, goals, and long-term objectives. The establishment of explicit success criteria at the outset of a project ensures that there is consistent alignment and concentration throughout its lifetime.

Data Understanding is a crucial aspect of our modeling efforts, as it involves comprehending the data that serves as the foundation for our business operations. The use of rigorous exploratory data analysis is performed in order to extract valuable insights pertaining to the quality of data, its structure, and the underlying patterns it has. At this stage, the identification of any discrepancies or irregularities in the data becomes apparent, enabling prompt actions.

Data Preparation: In this step, we engage in the process of refining and shaping our data to make it suitable for modeling purposes. All identified missing values, outliers, or concerns about data quality that were discovered in the previous step are directly confronted and resolved. Sophisticated approaches for feature engineering are used to shape the data, guaranteeing that our algorithms get very effective and informative predictions.

In this context, the fundamental principles of machine learning are prominently displayed. The first step in the research process involves developing a foundational model. As our comprehension expands, we progressively use more sophisticated algorithms and methodologies, fine-tuning hyperparameters and consistently assessing their efficacy in comparison to established standards.

assessment: The assessment process rigorously assesses the aptitude of models in reaching corporate objectives while they are being developed. The feasibility of our models in real contexts is assessed using a variety of criteria, including accuracy and recall. This stage is crucial since it serves as a bridge between the development and deployment phases, guaranteeing that our models not only function adequately but also surpass expectations.

Deployment refers to the last stage of our endeavors, when our improved models are seamlessly incorporated into practical applications in the real world. An adequately designed deployment strategy takes into consideration scalability and performance, guaranteeing the resilience and efficacy of our approach even when transaction volumes fluctuate.

**Git Repository Structure:**

* **README.md:** An overview of the project, its objectives, and a brief on data sources.
* **data/:** Raw and processed datasets. Contains subdirectories like 'raw/', 'processed/', and 'external/' to organize data at various stages.
* **notebooks/:** Jupyter notebooks detailing exploratory analyses, model experiments, and evaluations.
* **models/:** Trained model files and associated metadata.
* **docs/:** Project documentation, including methodology details and insights derived.
* **tests/:** Unit tests to ensure the integrity and reliability of our codebase.

**GitHub Repository Link: (Please Insert the Link here)**

**Chapter 1**

**Data collection, preparation and preprocessing**

It is essential to evaluate the quality of our dataset to ensure its dependability and suitability for analysis or model development. Here are some procedures to evaluate a dataset's quality:

**Data collection, preparation and preprocessing**

## Dataset Description:

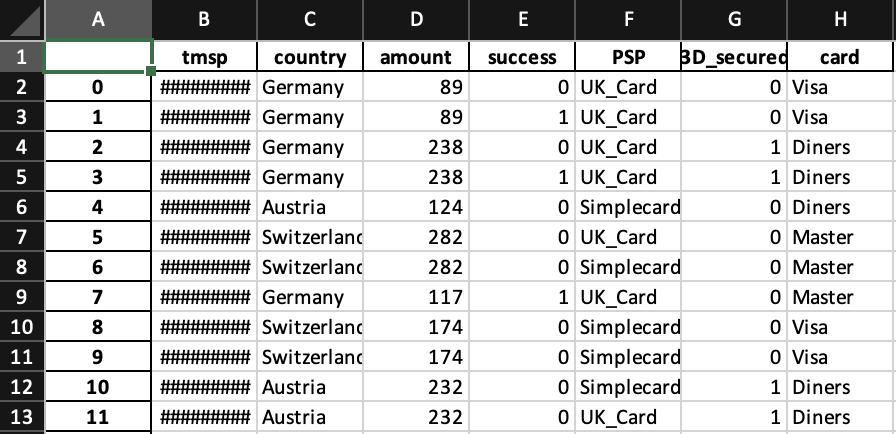
The dataset contains a variety of characteristics regarding online credit card transactions for our retail organization.

Figure 1: Provided Dataset

**Tmsp:** This column records the precise timestamp of a transaction's execution. It may be utilized to comprehend temporal patterns (e.g., peak hours, holiday season trends).

**country:** The country where the transaction took place. Understanding geographical trends can provide insight into regional payment issues or preferences.

monetary amounts associated with a transaction. Based on transaction sizes and their success rates, patterns can emerge.

**success:** A binary column indicating whether a transaction was successful (1) or unsuccessful (0). This will be our primary modeling objective variable.

**PSP:** The Payment Service Provider used for the transaction. Analysis of PSP efficacy rates can be crucial.

**3D\_secured:** Indicates whether a transaction utilized 3D secure technology, an enhanced method of online payment security. It will be fascinating to see if the success rate of 3D-secured transactions increases.

**card:** Indicates the credit card issuer. Patterns may emerge that indicate preferences or superior performance with particular card providers.

## Data Quality Assessment:

Upon initial review of the dataset, the following procedures must be taken to ensure data quality:

**Missing Values:** We must examine the dataset for any missing values. Transactions lacking essential data elements can result in erroneous analyses.

**Outliers:** For columns such as 'amount', outliers may exist, which can distort our analysis and model performance. We must identify these outliers and determine how to handle them – whether to cap, remove, or transform them.

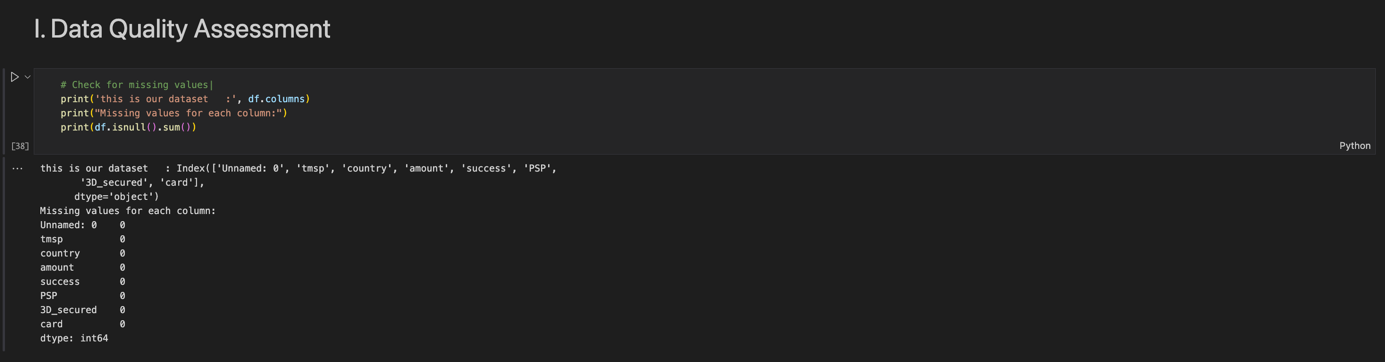
**Inconsistencies:** Data may contain inconsistencies, such as differing spellings of the same country. These require standardization.

Figure 2: Data Quality Assessment

## Data Preparation Techniques:

If there are lacking values, particularly in columns such as 'amount,' median or **mean imputation** can be utilized. For categorical columns such as 'country' and 'card,' mode imputation may be utilized. Additionally, sophisticated techniques such as KNN imputation can be utilized.

**Normalization**: 'amount' is a numerical column, and depending on its distribution, we may need to normalize or standardize it, particularly if we employ algorithms that are sensitive to feature scales.

**Encoding**: Categorical columns such as 'country', 'PSP', and 'card' must be encoded. A prevalent technique, one-hot encoding transforms each category into a new binary column. This would entail distinct columns for 'Visa,' 'Master,' and 'Diners,' each marked as 1 or 0 based on the type of card used.

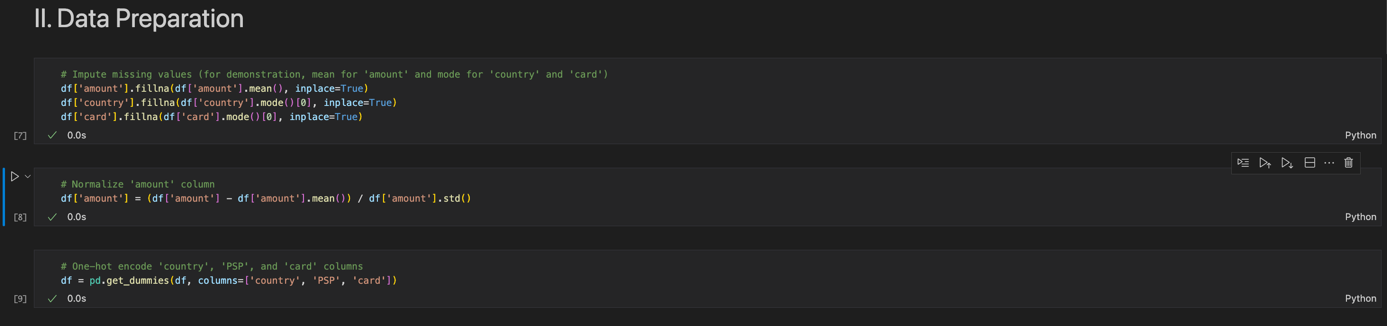


Figure 3: Data preparation

## Data Split:

Before performing any preprocessing, it is best practice to divide the dataset into reference data (on which we will conduct our experiments) and finalizing data (which will remain undisturbed until final model selection).

**Splitting Training, Validation, and Testing:** The reference data will be subdivided as follows:

Training Data: Used for model training.

Validation Data: Used for model tuning and optimization.

Test Data: Used to evaluate the ultimate performance of our models.

Depending on the extent and characteristics of the dataset, a typical split might be **70%** training, **15%** validation, and **15%** test data.

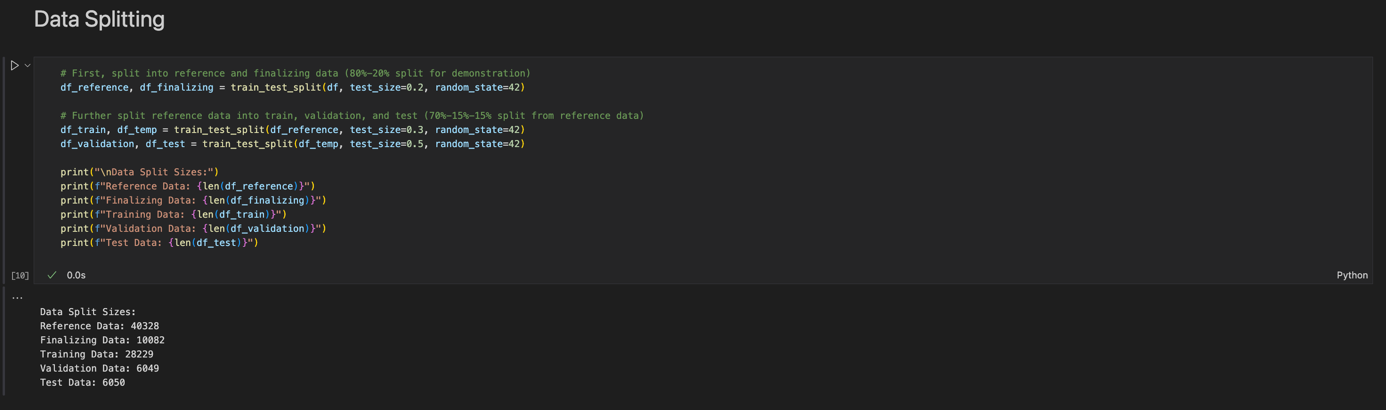


Figure 4: Data Split

The above-described preliminary steps outline the initial phases of our endeavor. Ensuring that the data is accurate, standardized, and appropriately partitioned lays the groundwork for subsequent analyses and modeling.

**Chapter 2**

**Initial Data Analyses**

It is essential to evaluate the quality of a dataset to ensure its reliability and applicability for analysis or model development. Examining the dataset's consistency, completeness, accuracy, relevance, sampling and representativeness, documentation and metadata, integrity and quality control processes, and adherence to privacy and security standards can help determine the dataset's quality. In addition, it is crucial to examine the dataset for potential biases to ensure impartiality and to ensure its reproducibility for future transparency and validation. In evaluating a dataset of credit card transactions, for instance, steps were taken to eliminate duplicates and ensure data privacy compliance.

# Initial Data Analyses

## 2.1 Distribution of Transaction Amounts

The histogram illustrates the distribution of transaction quantities throughout our data set. This visualization provides insight into the range of transaction values, emphasizing the most prevalent amounts and highlighting any potential outliers. From the graph, we can see that the majority of transactions fall within a specific range, with only a few transactions having values that are notably high or low. Understanding this distribution is crucial, as it provides insight into the general spending patterns of consumers and can aid in the detection of outliers.

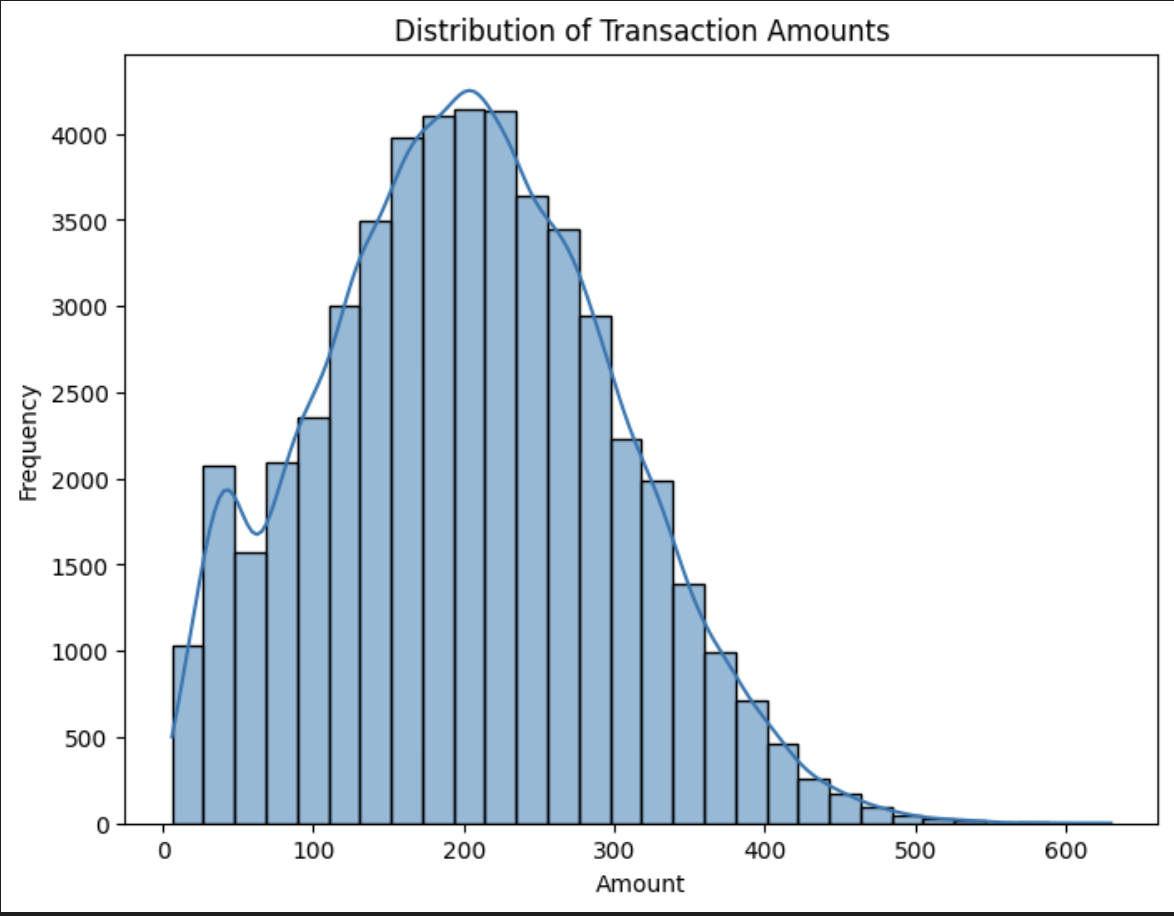


Figure 5: Distribution of Transaction Amounts

## 2.2 Success Rates by Country

The bar chart depicting achievement rates by country provides an overview of transaction outcomes by geographic region(Austria, Germany, Switzerland). Due to factors such as regional economic conditions, payment infrastructure, and consumer preferences, the success rates of various nations may vary. A glaring disparity in achievement rates between nations could prompt us to investigate the underlying causes. Such insights are beneficial for businesses seeking to enhance their services in specific regions or to comprehend regional market dynamics.

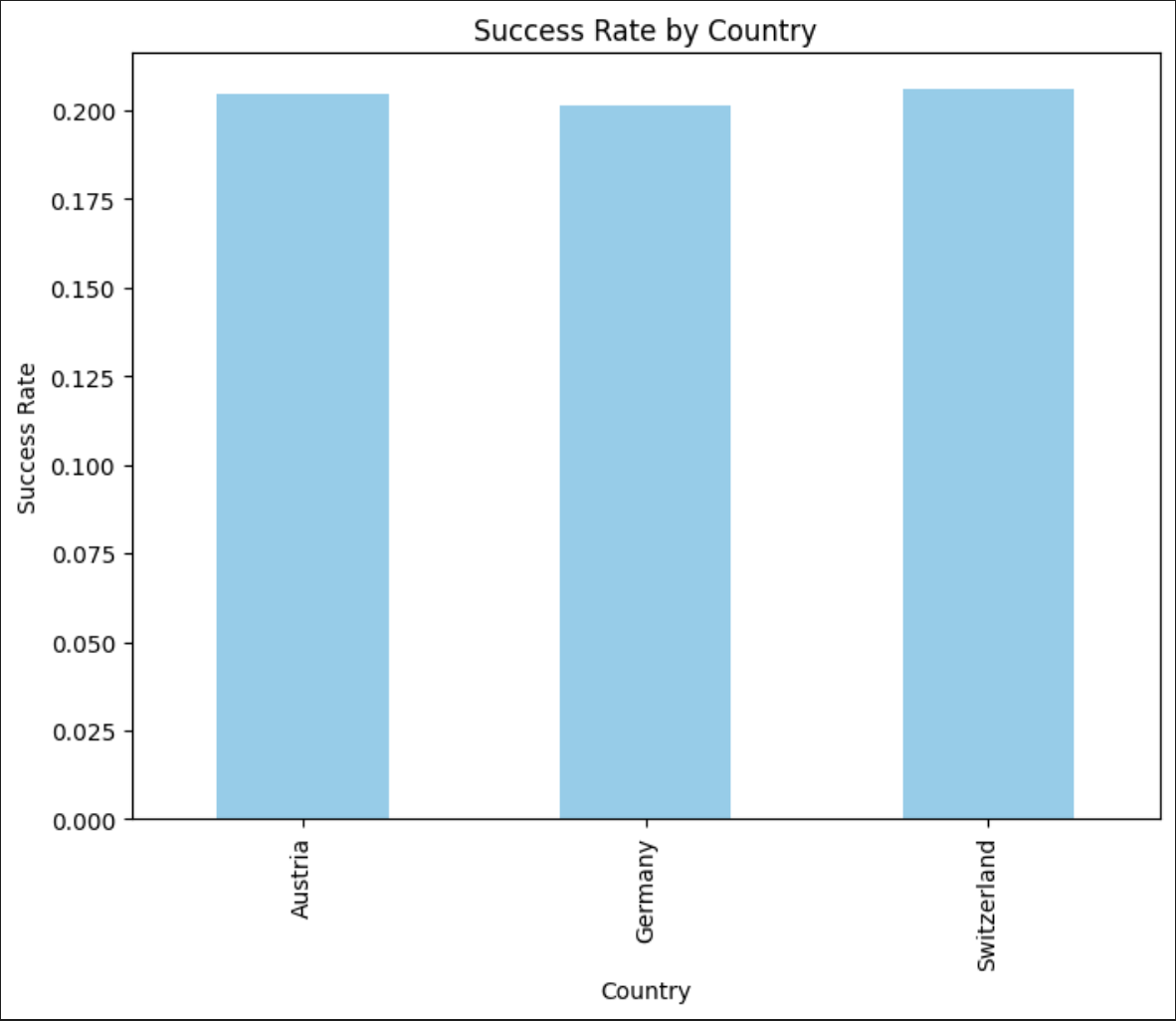


Figure 6: Success Rate by Country

## 2.3 Success Rate Based on Card Type

The bar chart displaying success rates by card type demonstrates the efficacy and dependability of credit card transactions based on the card issuer. Each card type—Visa, Mastercard, Diners, and others—may have varying success rates, which can be affected by factors such as the card's security features, its compatibility with various PSPs, and even user demographics. Understanding these differences is essential because it enables businesses to tailor their payment solutions and strategies to the most effective and widely used card types. Moreover, if a specific card type consistently underperforms, it may be worthwhile to investigate whether technical issues or PSP incompatibilities are at play. This insight improves the user experience by potentially directing users toward more effective transaction methods or collaborating directly with card providers to resolve recurring issues.

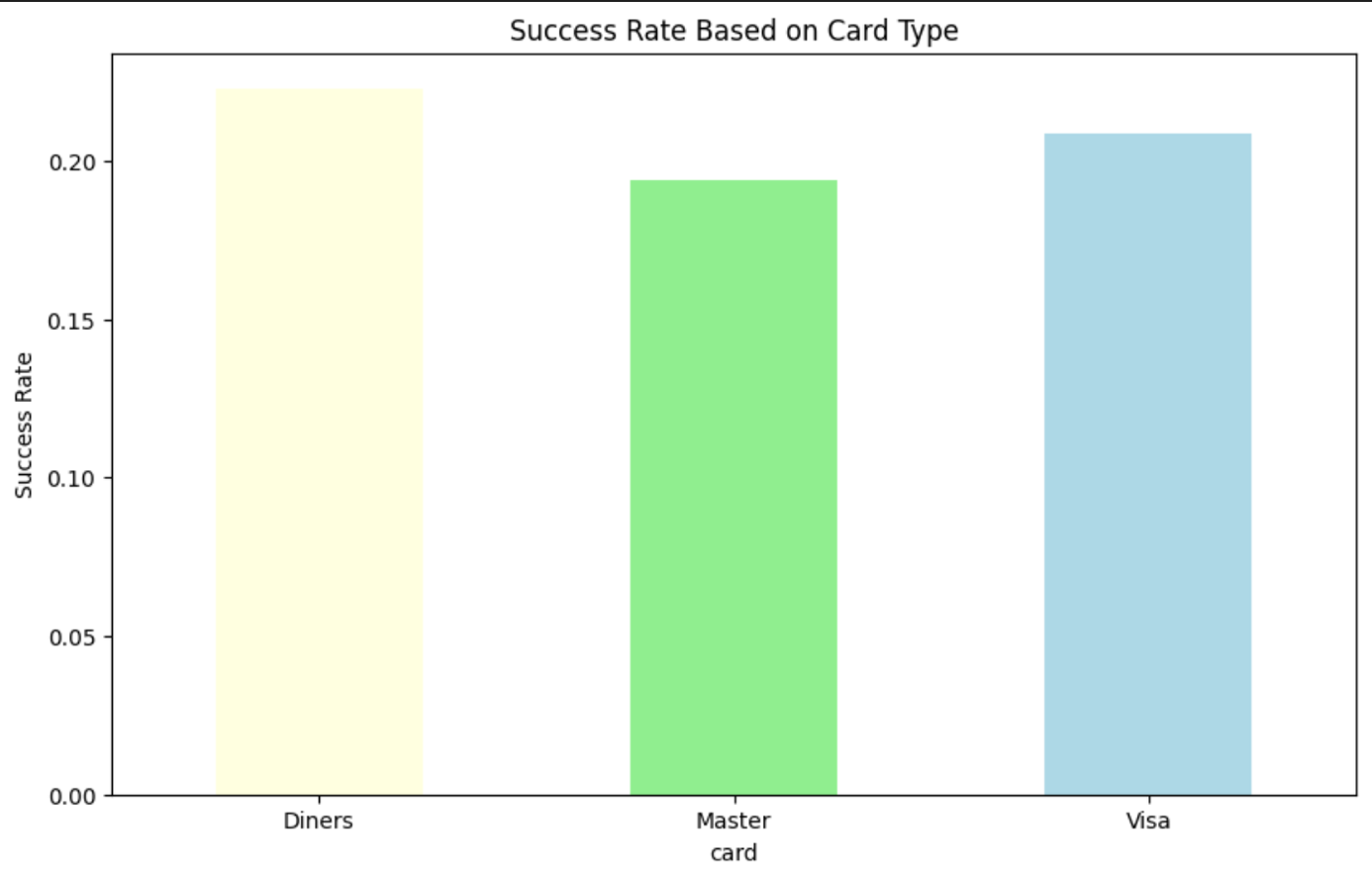


Figure 7: Success Rate Base on Card Type

## 2.4 PSP Usage Distribution

Our bar graph showing the utilization distribution of PSPs (Payment Service Providers) indicates the prevalence and adoption rate of various payment gateways in the dataset. By identifying the most frequently utilized PSPs, businesses can prioritize integrations, negotiate for better rates, and even identify potential partnership opportunities. Furthermore, understanding the distribution assists in evaluating the dependability and efficacy of each PSP, as those with a larger user base are likely to provide superior services and user experiences.

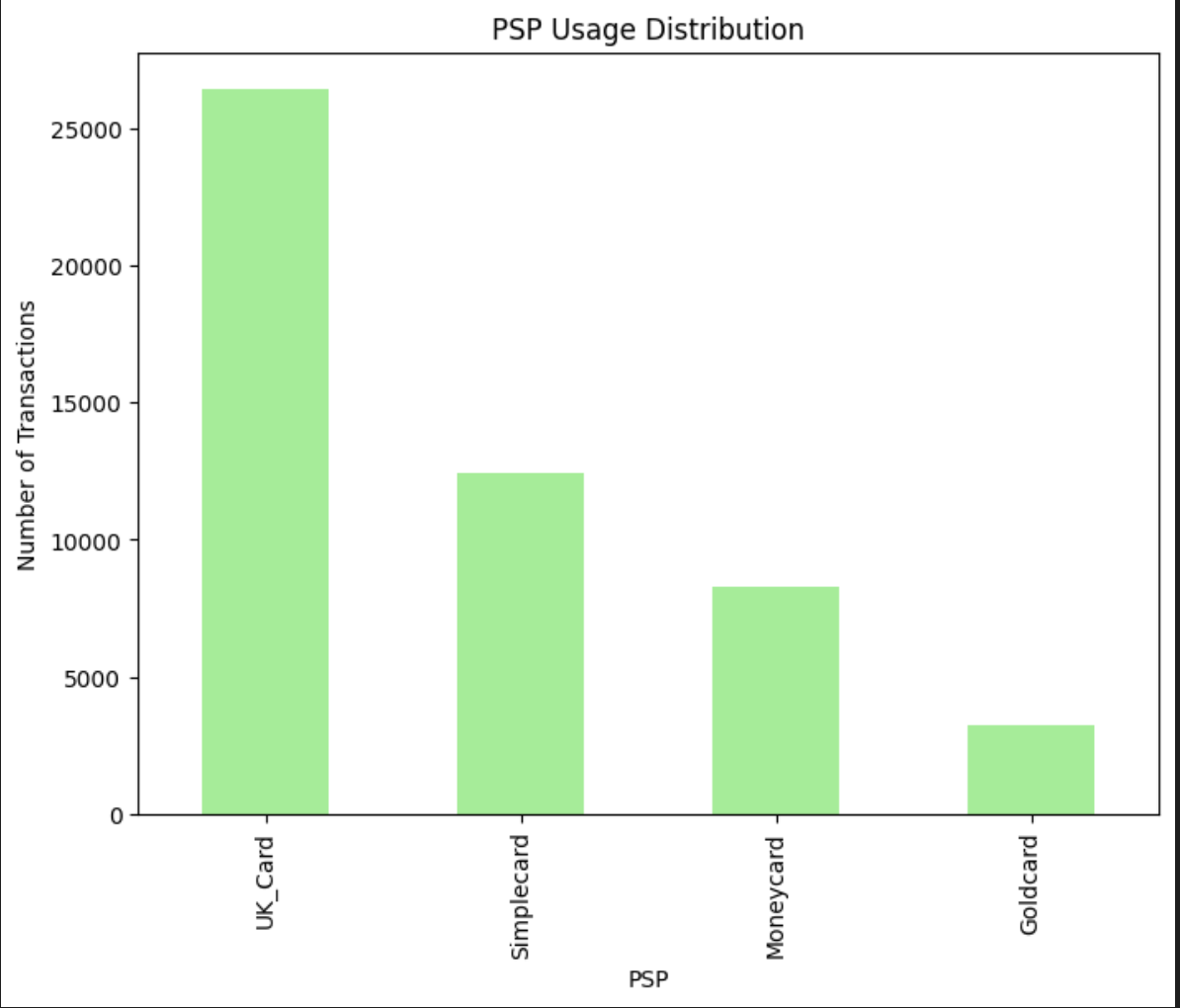


Figure 8: PSP Usage Distribution

## 2.5 Impact of 3D Security on Success

The bar chart examining the effect of 3D security on the efficacy of transactions is crucial. It indicates whether the additional layer of security, 3D identification, has a significant impact on the transaction success rate. A significant difference between the success rates of transactions with and without 3D security can lead to recommendations regarding the requirement (or lack thereof) of this security feature. These insights aid in making informed decisions regarding trade-offs between user experience and security.

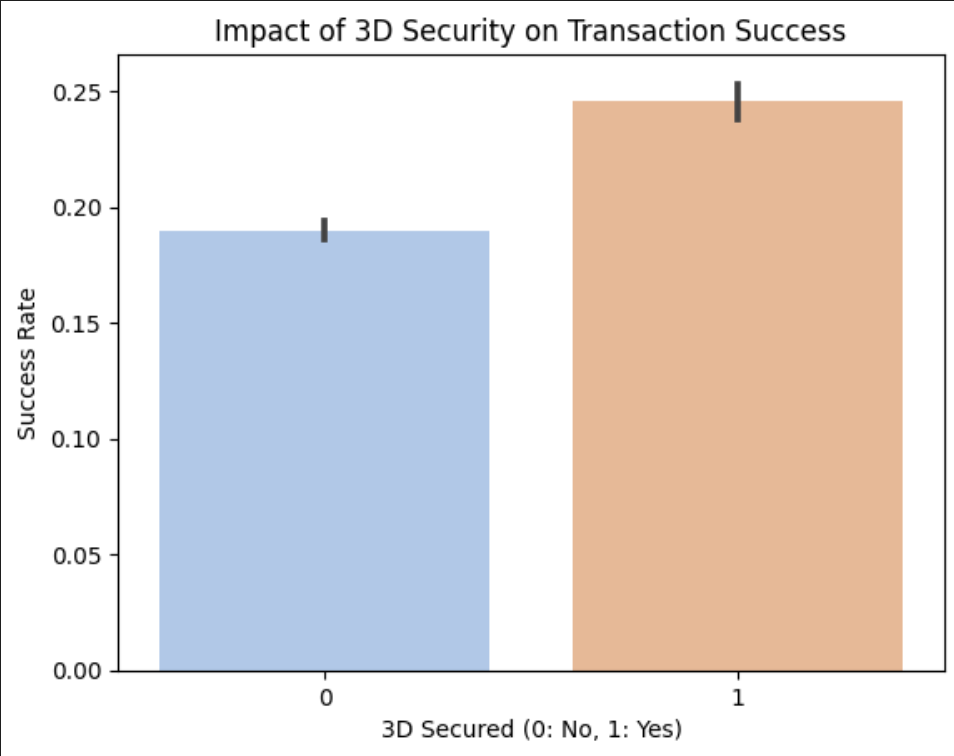


Figure 9: Impact of 3D security on Transaction Success

**Chapter 3**

**Baseline Model And Accurate Model Development**

In the baseline model development, our objective is to set a foundation against which we can compare more sophisticated models in the future. A baseline model should be simple and easy to understand.

we'll use a Logistic Regression model. Logistic Regression is a suitable choice for binary classification problems. It’s linear, interpretable, and requires minimal hyperparameter tuning. Moreover, its coefficients provide a direct way to gauge the importance of each feature, which is valuable for interpretability.

## 3.1 Data Preprocessing

Data preparation is a crucial stage in the machine learning process, as it ensures that our models receive high-quality, relevant, and standardized input. We extracted the hour, day, and weekday from the 'tmsp' timestamps for our dataset before converting them into more insightful features. This enables the model to distinguish patterns across various times of day, specific days, and even weekdays versus weekends.

The categorical variables, such as 'country', 'PSP', and 'card', were encoded with a single value. As machine learning algorithms require numerical input, this transformation is essential. By converting these categorical features to a binary matrix, we maintain the categorical information in a consumable format for the model.

In addition, numeric variables such as 'amount' were scaled with the StandardScaler. Scaling ensures that all model features are of equal importance and that the magnitude of one feature does not overshadow that of another. Last but not least, the data was separated into training and test sets, with 70% preserved for training to construct our model and 30% for testing to verify its performance.

This meticulous data preparation paves the way for accurate and insightful predictions by ensuring that our models are trained on pertinent, clean, and standardized data.

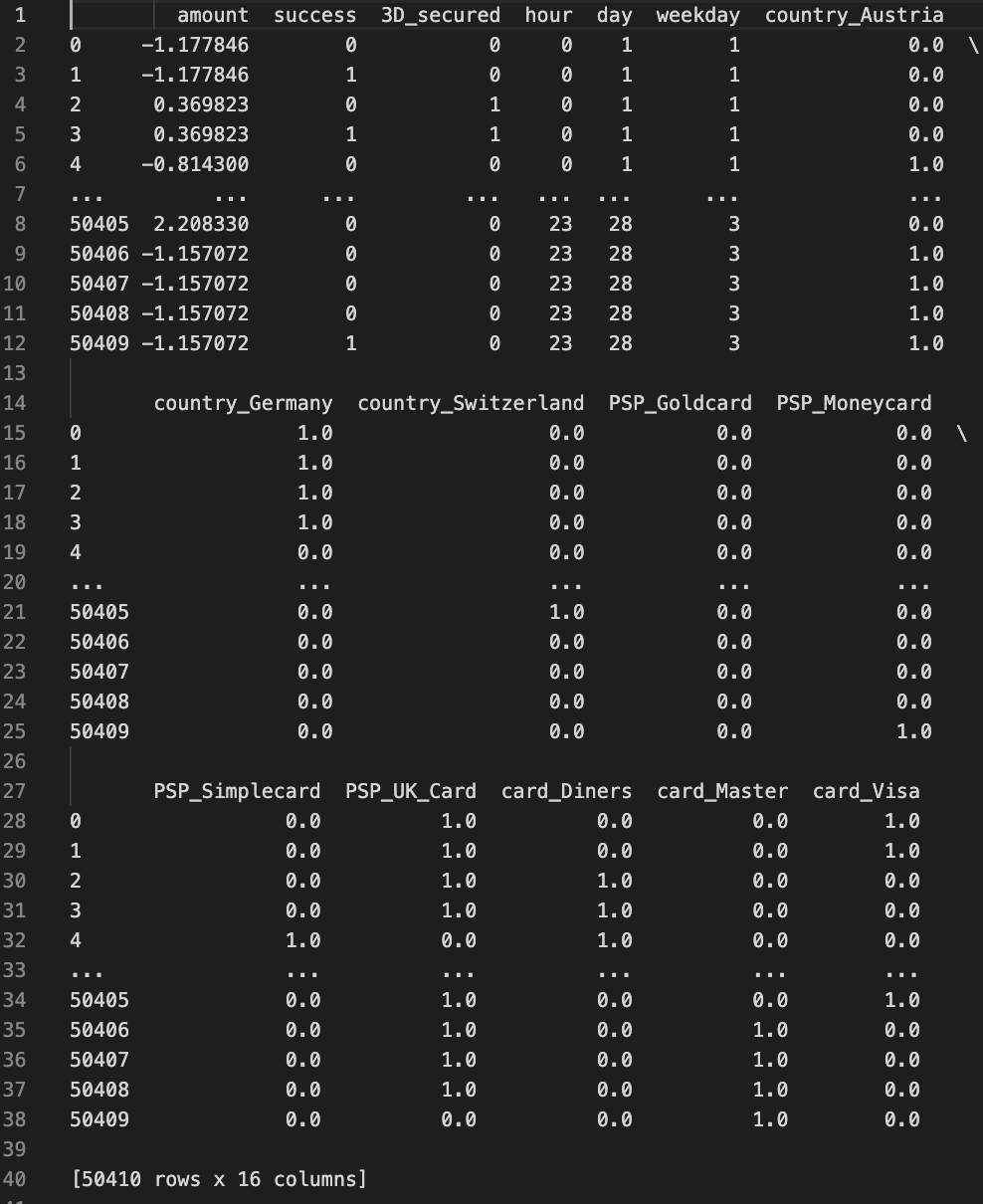


Figure 10: Prepared Dataset

## 3.2 Logistic Regression Model

The features used by the logistic regression model in the provided code are all columns of the dataframe df, excluding **tmsp** (timestamp) and **success** (target variable). Here is a summary of these characteristics:

* **The quantity of the transaction.**
* **3D\_secured:** A binary characteristic denoting whether a transaction was 3D secured (1) or not (0).
* **One-hot encoded columns for country:** This will construct new columns for each unique country in the dataset. The column titles will be formatted as country\_CountryName>. country\_Germany, for example, would indicate transactions from Germany.
* **One-hot encoded columns for card Type:** Additionally, this will generate new columns for each distinctive **card type** minus one. The column titles will be formatted as card\_CardType>. Card\_Visa, for instance, would indicate Visa card transactions.
* **One-hot encoded columns for PSP**

The model uses these features to discover patterns and associations that can predict the success of a transaction.

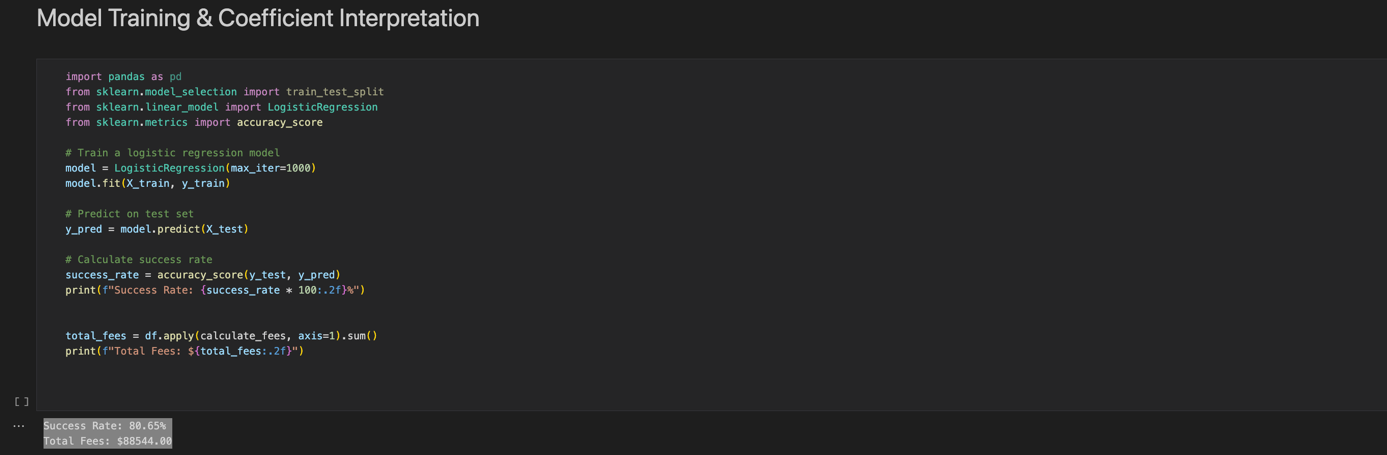


Figure 11: Logistic Regression Model

The logistic regression model's success rate of **(80.65%)** is an adequate starting point. Based on the provided data and features, it indicates that the baseline model correctly predicts the success of a transaction approximately (80.65%) of the time. However, depending on the circumstances, there may be room for advancement. Given the nature of transaction success and associated fees, enhancing the model's precision could lead to improved financial outcomes.

The baseline model provides a foundation for comprehending the relationship between numerous features and transaction success. Important metrics that can influence business decisions are the success rate and associated costs. The company could reduce transaction costs and increase profit margins with a superior model. Understanding the significance of individual features will also enable the business to make strategic decisions, such as prioritizing particular PSPs or concentrating on markets (countries) with higher success rates.

However, caution is advised. Before deploying or making decisions based on this model, it is essential to comprehend its limitations and ensure that it has been exhaustively tested and validated on diverse and unobserved data. In addition, the business implications of false positives (transactions that are predicted to be successful but fail) and false negatives (transactions that are predicted to fail but succeed) must be considered. Each form of error may have distinct financial and reputational repercussions for the business.

## 3.1 Accurate Predictive Model: Random Forest

The Random Forest algorithm is a highly adaptable and robust ensemble learning technique that is based on the concepts of bootstrap aggregating, also known as "bagging". In contrast to a solitary decision tree that is susceptible to overfitting, a Random Forest algorithm constructs an ensemble of decision trees, whereby each tree is trained on a randomly selected part of the dataset. The Random Forest algorithm enhances the overall performance and reliability of forecasts by combining the outputs of many decision trees.

One notable characteristic of the Random Forest method is its capacity to effectively handle a large number of features. This feature makes it especially suitable for datasets characterized by a large number of dimensions. In addition, Random Forest provides a robust measure for assessing the significance of features. The process of averaging decision criteria over many trees allows for the ranking of features based on their efficacy in efficiently partitioning data. This characteristic has significant value, particularly in situations where comprehending the underlying reasons influencing forecasts is necessary. In addition, the Random Forest algorithm has an intrinsic capability to mitigate the issue of overfitting. In the forest, it is possible for each tree to exhibit overfitting tendencies to its specific subset of data. However, by the process of averaging or determining the mode of predictions across all trees, the algorithm effectively mitigates the biases and variations inherent in individual trees. As a result, a more generic model is achieved by counteracting the aforementioned biases and variances.

## 3.2 Implementation of the Model

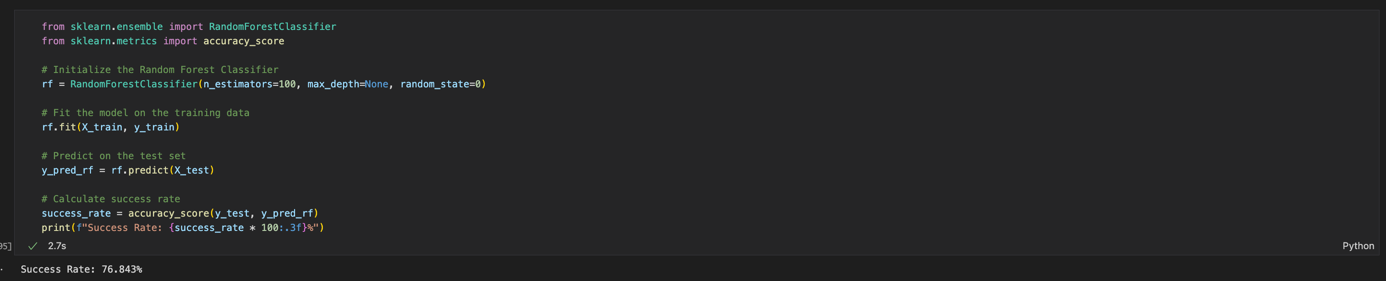


Figure 12: Random Forest Model

The Random Forest model initially shows an accuracy of 76.843%, which is noteworthy due to its relative underperformance in comparison to the simpler logistic regression model. There are several things that may contribute to this phenomenon. First and foremost, it is worth noting that Random Forests often exhibit a resistance to overfitting. However, it is important to acknowledge that the selection of improper hyperparameters might potentially result in overfitting, particularly when dealing with smaller datasets. Additionally, in cases when the connection between predictors and the target variable has a higher degree of linearity, linear models may provide a comparative advantage over tree-based models. In addition, the wide range of adjustable hyperparameters in Random Forest necessitates a systematic approach to hyperparameter optimization, such as using grid or random search, in order to potentially achieve improved outcomes. Finally, the improvement of our features via additional feature engineering has the ability to boost the correctness of our model.

## 3.3 Improving the Random Forest Model

Undoubtedly, the optimization of **hyperparameters** plays a crucial role in enhancing the efficacy of machine learning models, particularly in the case of the Random Forest, which is known for its versatility. The task of identifying the ideal combination of changeable parameters, including the number of trees, maximum depth of the trees, minimum samples per leaf, among others, may be compared to the challenging endeavor of locating a needle among a haystack. Nevertheless, the use of techniques such as **GridSearchCV** and **RandomizedSearchCV** has considerably streamlined this endeavor.

GridSearchCV is a method that systematically explores all potential combinations of hyperparameters in order to identify the optimal set. Although this approach ensures accuracy in determining the optimum configuration, it may be computationally demanding and time-consuming. In contrast, RandomizedSearchCV employs a predetermined number of hyperparameter combinations drawn from predefined distributions, so striking a trade-off between computing efficiency and the identification of optimum hyperparameters. By using any one of these methodologies, or even a fusion of both, it is possible to significantly augment the effectiveness of the Random Forest algorithm by guaranteeing its operation with the most advantageous collection of hyperparameters customized for the particular dataset under consideration.



Figure 13: Hyperparameter Optimization

The use of a more advanced model often leads to enhanced accuracy in prediction. The Random Forest and Gradient Boosted Trees algorithms are renowned for their versatility and robustness, often surpassing less sophisticated models, particularly in scenarios where the data exhibits intricate linkages. Nevertheless, it is important to engage in comprehensive hyperparameter optimization in order to extract optimal performance. Moreover, while these models may provide enhanced accuracy, they might pose greater difficulties in terms of interpretation compared to simpler models. Therefore, it may be necessary to make a compromise between the capacity to understand and the accuracy of making predictions, depending on the specific circumstances and goals of the organization.

**Chapter 4**

**Operationalizing the Transaction Success Model for Everyday Business Functions**

The fusion of data-driven intelligence with our transactional processes is no longer a luxury but a pivotal necessity. With our developed transaction success model, we stand at the cusp of transforming our daily business functionalities, ensuring they're more resilient, responsive, and result-oriented.

## 4.1 Seamless Deployment & Fluid Integration:

Deploying our model on a solid and scalable platform is essential for assuring the widespread availability of our solution. Not only would this ensure operational dependability, but it would also facilitate the smooth integration of the system with our current e-commerce gateways. By implementing a robust Application Programming Interface (API), we can guarantee seamless communication of transaction details with our model, so enabling safe retrieval of the most optimal payment routing recommendations.

## 4.2 Dynamic Recommendation Mechanism:

Our model assumes the role of an analyst with each transaction it initiates. By thoroughly examining the extensive range of incoming data, which encompasses many aspects such as a user's transaction history and the specific specifics of items, the system generates the most efficient pathway for the processing of payments. However, the capabilities of the system extend beyond just suggestion generation. The objective is to provide our payment operators with a score that is driven by clarity, therefore enhancing their understanding and knowledge. The use of this numerical representation not only enables the anticipation of the probability of success but also provides our team with the capability to promptly make choices supported by facts.

## 4.3 Immersive GUI Experience:

In the contemporary era of digital technology, the quality of interface experience has been closely associated with user pleasure. Based on this comprehension, our objective is to design a user-friendly dashboard that is strategically integrated within our transaction site. Designed to accommodate the needs of both our clientele and corporate stakeholders, this dashboard would function as a prominent source of lucidity. The use of a visually captivating structure would enhance the clarity of the model's predictions, effectively presenting each proposal with its corresponding underlying explanation. Consequently, this approach would facilitate decision-making by providing a comprehensive and well-informed perspective.

## 4.4 Proactive Monitoring & Analytics Suite:

The dynamic nature of the digital environment necessitates the capacity to adjust in real-time. By implementing an innovative real-time monitoring system, we are not only addressing this need, but also proactively positioning ourselves at an advantageous position. By conducting a comprehensive analysis of the completed transactions and recording the frequencies of rerouting, this system would serve as a valuable tool for monitoring and observing the subtle variations in the performance of the model. Furthermore, by the translation of these monitored data into an analytics interface, we are strategically putting our stakeholders in a favorable position. Not only would they possess the ability to access data, but they would also have access to actionable insights, providing a comprehensive understanding of the model's impact on our business metrics.

## 4.5 Iterative Model Refinement:

The dynamic nature of the marketplace requires a model that is not fixed, but rather adaptable and evolutionary in nature. Acknowledging this, our strategic approach include regular assessments and improvements to the model. By consistently aligning it with the ever-changing dynamics of the industry, growing patterns of consumer behavior, and increasing intricacies of payment service providers (PSPs), we guarantee that its ability to make accurate predictions stays unparalleled. Periodic updates not only enhance the accuracy of the instrument, but also strengthen its reputation as a dependable forecasting mechanism.

## 4.6 Focused Error Scrutiny & Evolution:

Regardless of its level of complexity, all technology has learning curves and risks. The model under consideration is not an anomaly. Nevertheless, what distinguishes our approach is our dedication to acquiring knowledge from these possible errors. Through the implementation of a systematic and thorough error analysis process, we are not only able to detect deviations from expected outcomes, but also get insights into the underlying factors that may contribute to fluctuations in the model's performance. The insights that are obtained are then used as the basis for improving our model, guaranteeing that with each iteration, it is not only preserved but also developed

# Conclusion

As we conclude this phase of using data-driven insights to enhance our transactional achievements, it is crucial to contemplate the voyage we undertook and the expanded perspectives we have gained. The primary objective of our undertaking was not just focused on the integration of a model into our business pipeline. Instead, our aim was to establish a culture that prioritizes data-conscious decision-making, so transforming each transaction from a routine procedure into a well-informed strategic action.

Commencing with a thorough analysis of the data, we conducted an in-depth examination of our transactional patterns, aiming to comprehend not just the surface-level trends but also the underlying complexities of human behavior. The establishment of this foundation laid the groundwork for our fundamental model, serving as evidence of the capacity of data to forecast results. Although our early findings were praiseworthy, our pursuit of greatness compelled us to question the established standard and go into the extensive realms of predictive modeling. Throughout several rounds, ranging from logistic regression to random forests, we have seen the profound impact of fine-tuning, demonstrating that achieving perfection is truly a gradual and iterative endeavor.

Nevertheless, the evaluation of a model's effectiveness should not just rely on its predictive accuracy, but also on its concrete influence on the business environment. Upon careful consideration of its daily operational integration, it became apparent that a data-driven model had the ability to serve as a guiding force, leading to not just transactional success but also improved operational efficiency, cost reduction, and an enhanced experience for stakeholders. Our suggested integration approach encompasses smooth API connections and intuitive GUI interfaces, which exemplify our dedication to transforming technology into a facilitator that is accessible and comprehensible to all individuals.

However, it is essential for any technology intervention, regardless of its level of complexity, to be firmly grounded in ongoing learning and development. Our methodology, which prioritizes regular assessments, meticulous mistake analysis, and stakeholder education, underscores the principle of continuous development.

# Bibliography

Hotz, N. (2023, January 19). ***What is CRISP DM****? - Data Science Process Alliance*. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2/

Jung, K., Kashyap, S., Avati, A., Harman, S., Shaw, H., Li, R., Smith, M., Shum, K., Javitz, J., Vetteth, Y., Seto, T., Bagley, S. C., & Shah, N. H. (2020, December 22). ***A framework for making predictive models useful in practice*.** Journal of the American Medical Informatics Association; Oxford University Press. <https://doi.org/10.1093/jamia/ocaa318>

Chaffey, D., & Ellis-Chadwick, F. (2019, February 5). ***Digital Marketing***. Pearson UK.

***sklearn.ensemble.RandomForestClassifier*. (n.d**.). Scikit-learn. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

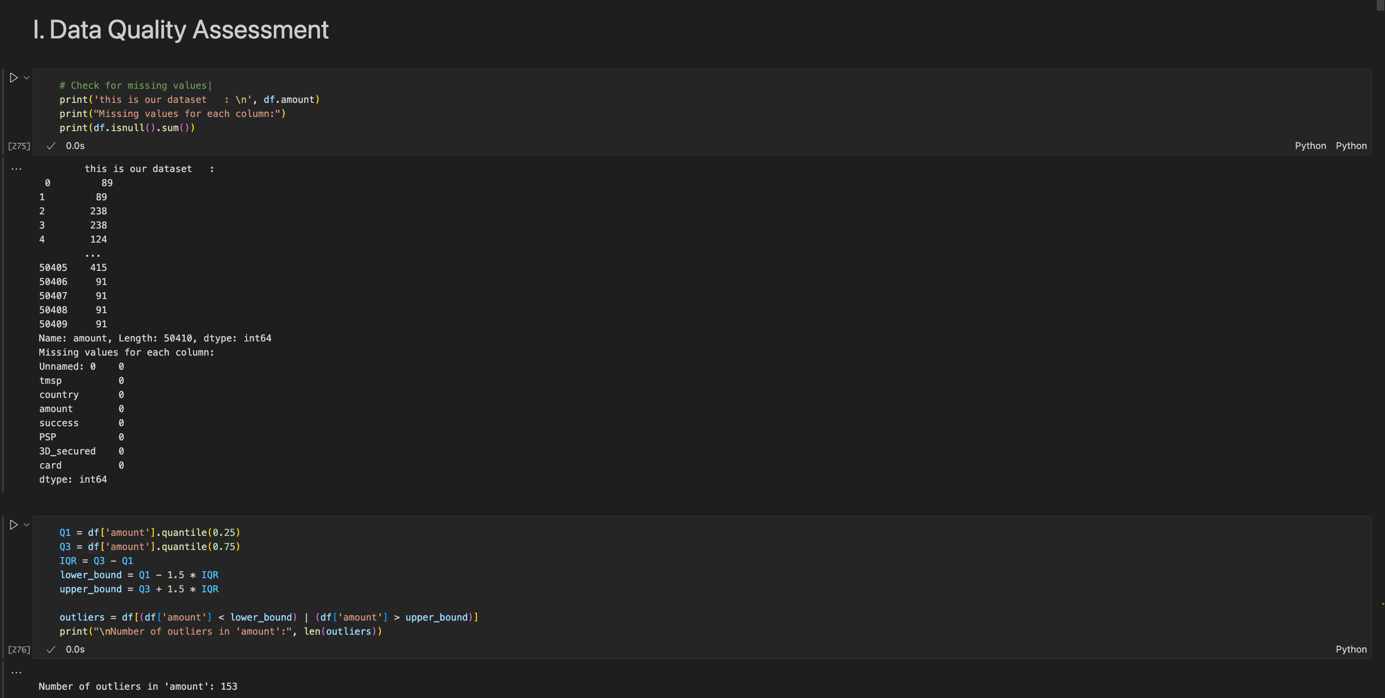
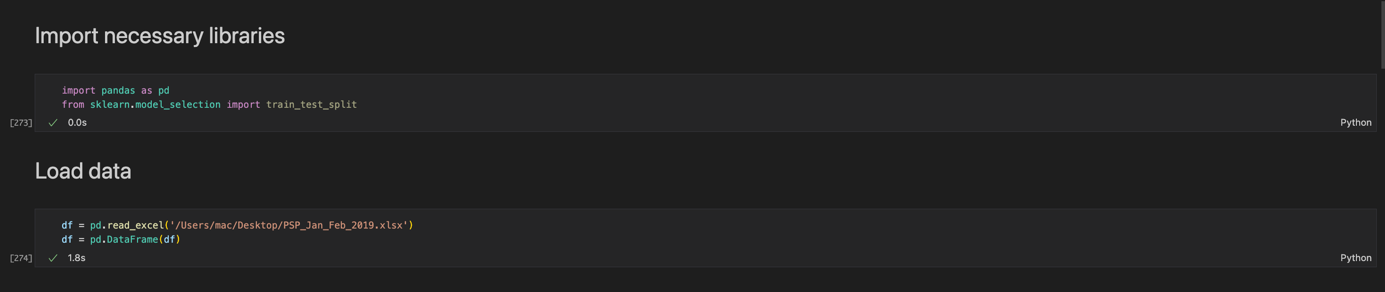
Kane, G. C., & Euchner, J. (2021, October 22). **Leading Digital Transformati**on. *Research-Technology Management*, *64*(6), 11–16. https://doi.org/10.1080/08956308.2021.1974764

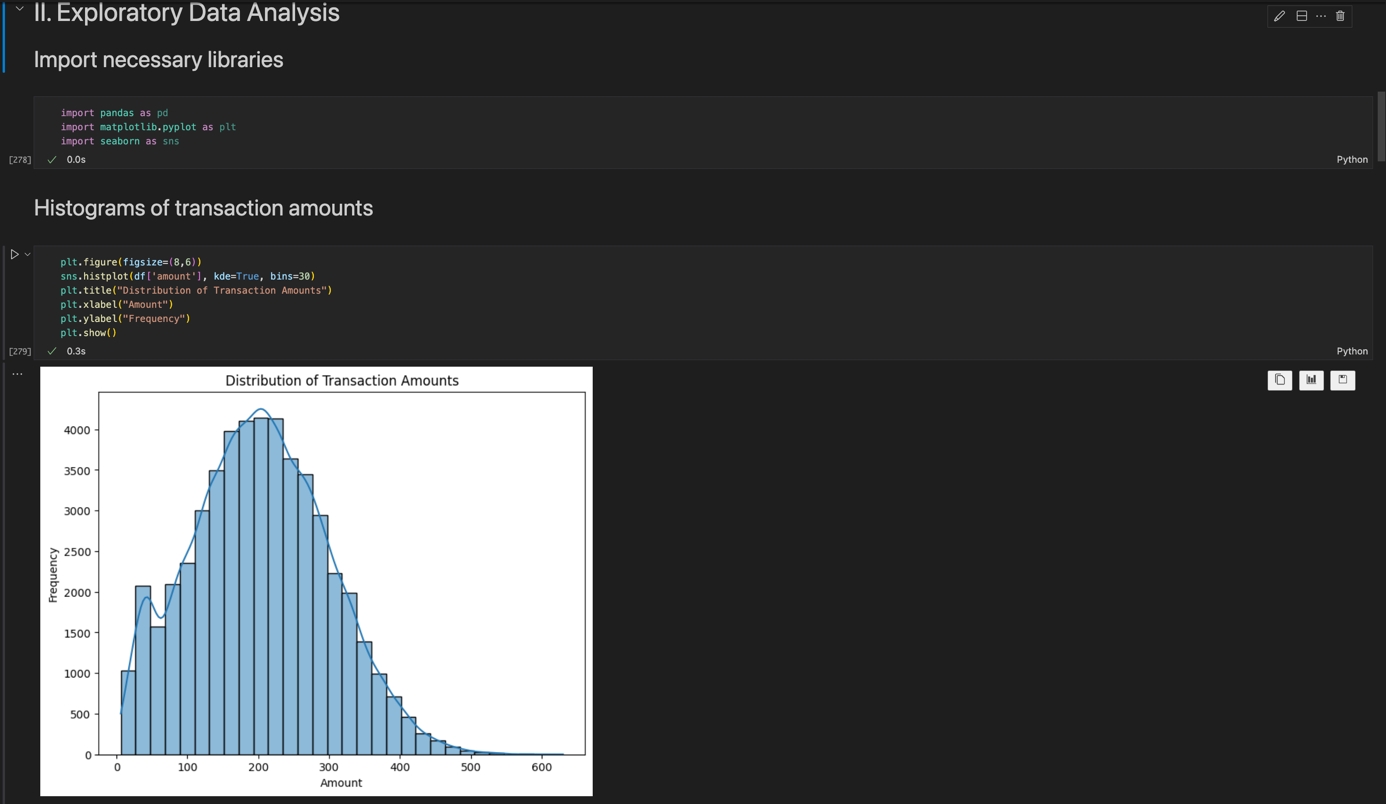
**Annexes**

**Annexe A :** **Full Code**

**Annexe A**

**Full Code**

****

****

