**BPP Business School**

# Coursework Cover Sheet

Please use this document as the cover sheet of for the 1st page of your assessment. Please complete the below table – the grey columns

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Module Name | Programming for Data Analysts (PDA) | | | |
| Programme Name | MSc Management with Data Analytics | | | |
| Student Reference Number (SRN) |  | BP0293328 |  | |
|  |
| Assessment Title |  | Loan data automation of Apex Financial Services CW3[S] | |  |

Please complete the yellow sections in the below declaration:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Declaration of Original Work:  I hereby declare that I have read and understood BPP’s regulations on plagiarism and that this is my original work, researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Business and Technology.  The word count, excluding contents table, bibliography and appendices, is 2500 words. | | | | |
| Student Reference Number | \_\_\_\_\_BP0293328\_\_\_\_ | Date: | 2024/07/01 |  |

By submitting this course work you agree to all rules and regulations of BPP regarding assessments and awards for programmes.

Please note that by submitting this assessment you are declaring that you are fit to sit this assessment.

BPP University reserves the right to use all submitted work for educational purposes and may request that work be published for a wider audience.

# MSc Management with Data Analytics

## SUMMATIVE Programming for Data Analysts

**Leveraging Data Analytics for Operational Excellence in Financial Services: A Case Study of AFS**

# Table of Contents

[**Coursework Cover Sheet** 1](#_Toc170839863)

[MSc Management with Data Analytics 3](#_Toc170839864)

[**SUMMATIVE** Programming for Data Analysts 3](#_Toc170839865)

[Table of Contents 4](#_Toc170839866)

[**Introduction** 5](#_Toc170839867)

[**Approach** 7](#_Toc170839868)

[**Methodical Approach to Data Implementation** 7](#_Toc170839869)

[**Data Fusion and Consolidation** 9](#_Toc170839870)

[**Data Scrubbing and Preprocessing** 10](#_Toc170839871)

[**Data Insight Exploration** 11](#_Toc170839872)

[**Recommendations for Future Enhancements** 14](#_Toc170839873)

[**Conclusion** 15](#_Toc170839874)

[**References** 16](#_Toc170839875)

[**Appendices** 17](#_Toc170839876)

[Apendix-1 17](#_Toc170839877)

[Apendix-2 24](#_Toc170839878)

# Introduction

This report details the strategic plan for Apex Financial Services (AFS) to improve operations and reduce risk through innovative programming solutions This report acts as an instruction guide in which AFS can use the right technology and data-driven insights for optimizing operational efficiency to stay ahead of other competitors. The report also starts with setting out its aim and scope, thus introducing the reader to important informational features on what elements will be explored further in this document. This is then compared to a deep dive into the current business environment that affects other corporations, including AFS (Rushchyshyn et al. 2017). This landscape knowledge is important because it helps AFS to understand the macro variables that will influence strategic implementation of proposed programming solution/s. By grounding the report in a deep dive into their current market situation and competitive forces, AFS can make sure those recommended solutions are bespoke to them but also in line with what its customers need and what's going on within financial services more broadly. In doing so, Europe wants to create a stepping stone for AFS's strategic roadmap and enable operations resilience while preventing risks as well as establishing a sustainable strategy in view of the dynamic business environment (Chanias et al. 2019).

This dataset is a considerably useful information-rich resource that we can use to actually make the kinds of changes at Apex Financial Services (AFS) you are hoping for. The dataset consists of a PDF with exhaustive business loan transaction history, and an Excel file containing sales data. Together, these complementary resources form an extensive and detailed fabric of historical context that AFS can use to power advanced data analytics as well as ultramodern supervised machine learning models. Using this dataset, AFS is able to access more extensive details on its lending activity and customer trends in addition to sales performance. Using advanced analytics and insights will allow the firm to spot areas of improvements, accelerate its approval process for lending products thus, driving overall operational efficiencies. Based on this analysis, organizations will develop new innovative programming solutions that are intelligent enough to perform what we call common routine tasks by freeing up resources who can help humans in delivering a better customer experience (Raguseo and Vitari 2018).

A center-piece of this extensive Report is a specific functionality design for addressing some high-value User stories identified in relation to the Apex Financial Services (AFS). The main purpose is to automate various kinds of repetitive tasks in the loan processing workflow; which are important parts of optimization operation work for organizations. With the resources freed up from these repetitive manual processes, AFS will be able to engage more deeply with its clients and personalize customer service, two key differentiators for a leading player in today's fiercely competitive financial services sector. Given the significance of this automation initiative, the report tackles challenges that are focussed around code development and maintenance with detailed attention (Hsu 2020). The use case highlights the advantage of using reusable code that will not only make it easy to scale but also help in adding extra functionalities based on evolving business requirements at AFS. Taking this approach as a business will allow the organization to stay nimble and respond quickly to market changes, setting it up for success in future. Additionally, the report outlines important regulatory and ethical aspects with respect to data privacy issues as well as compliance that every financial organization or company needs to take into account. As AFS uses innovative programming offerings to streamline its operations, it also must ensure that these solutions meet the highest level of data stewardship and protection in order not only support but build upon trust with customers and other stakeholders. The report offers a thorough structure to accommodate these challenges, making certain any solutions are industry compliant and adhere to best practice from an ethical perspective Summary: This report creates a strategic road map that will help AFS to exploit this new programming-driven transformation, and perfect the art of being truly competitive in today's fast-moving world finance landscape by meeting these core user requirements with sound code development practices while considering regulatory/ethical standards (Weaver and Treviño 1999).

Understanding the need for reform, this paper highlights to AFS both the consequences of inactivity and what can be achieved if it implements these proposed improvements. If this revolution of the cords is not adopted for programming needs, we could lead to inefficiencies which result in operational costs increase and quality service about to drop, it a no-win situation as far efficiency reach will be concerned leading causing our customers dry their trust on us eroding AFS responsible position amidst an intense fight between player inside financial services sector. To manage these risks and realize the full potential of the innovations proposed we recommend strategic investment in programming and data analytics skilled human resources. These unique competencies will be critical to bring in the required innovation and support AFS's journey towards organic scaling considering its volatile market environment. By empowering its workforce with cutting-edge technical capabilities, AFS can transform in this way into an industry leader that continuously adjusts based on the desires of customers and advances within the market by rapidly deploying innovative data-driven programming solutions. The report concludes by setting the stage to effectively build on these findings, reiterating that AFS needs a clear answer to those same three questions, what are the risks of inaction, how can it leverage technical expertise and where might redoubling their commitment deliver breakthrough innovation given its dynamic new information services-centric environment? By following these recommendations, AFS will be well on its way to an improved position regarding their operational resilience going forward so they can improve customer-facing experiences and remain competitive years into the future (Baccarini et al. 2004).

Approach

## Methodical Approach to Data Implementation

The approach taken to implement the solution for Apex Financial Services (AFS) was underpinned by a methodical data analysis framework, tailored to enhance operational efficiencies and mitigate risks inherent in manual loan processing.

**Requirement Gathering and Data Acquisition**

Following an adoptive regression, a more comprehensive set of consultations took place with the loan team (user), marketing department and compliance to define crisp user requirements. With these insights in mind, we then went forward with the selection and acquisition of two main datasets - a PDF file containing extensive historical loan records and an Excel document documenting current sales data. Based on metadata standards, they performed validations to check that both datasets are relevant and complete before allocating them for further analyses (Hao et al. 2023).

**Data Preparation and Quality Assurance**

A meticulous data preprocessing phase followed, focusing on cleansing operations such as handling missing values, resolving data inconsistencies, and eliminating duplicates. Special attention was devoted to the Excel file to mitigate the risks associated with shared access and potential data integrity issues (Zhu et al. 2007).

**Exploratory Data Analysis (EDA)**

It is observed that the patterns and insights obtained from the datasets were carried out through EDA. Descriptive statistics, correlation analysis and visualization in the form of histogram scatter plots were used to explore the various loan attributes, borrower demographics and approval trends. This step gave a basic understanding which helped me to model in future steps.

**Model Development and Validation**

The solution revolved around recommending supervised machine learning in order to predict the outcome of a loan on important data. We explored various algorithms such as logistic regression, decision trees and ensemble methods followed by a significant amount of parameter tuning along with validation techniques. Cross-validation will ensure model performance over a variety of loan characteristics (Orji et al. 2022).

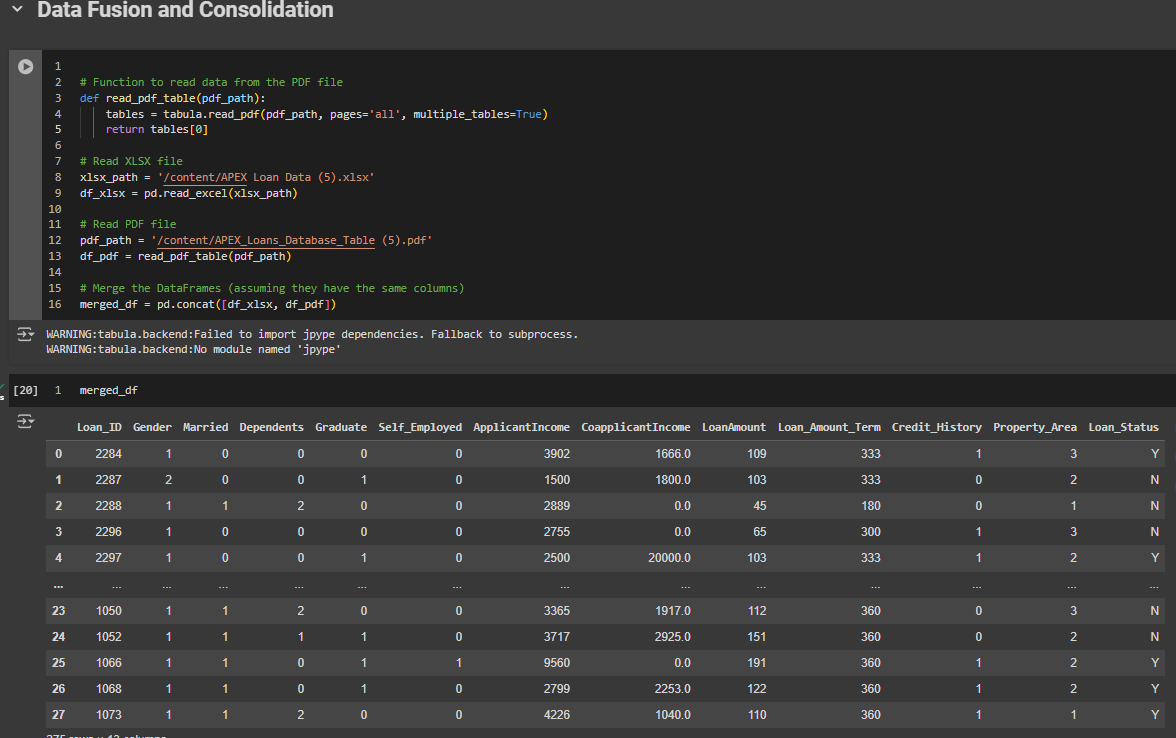
**Integration and Deployment**

These models were then recommended to be integrated within an automated loan processing system that was capable of routing through the approvals and reducing manual intervention. The implementation will be mainly done using Python, and used the languages like Pandas, Scikit-learn & Matplot-lib packages for its integration as well deployment. The iterative process of gathering continuous feedback from end-users and streamlining it with operational objectives was further enhanced.

**Monitoring and Sustainability**

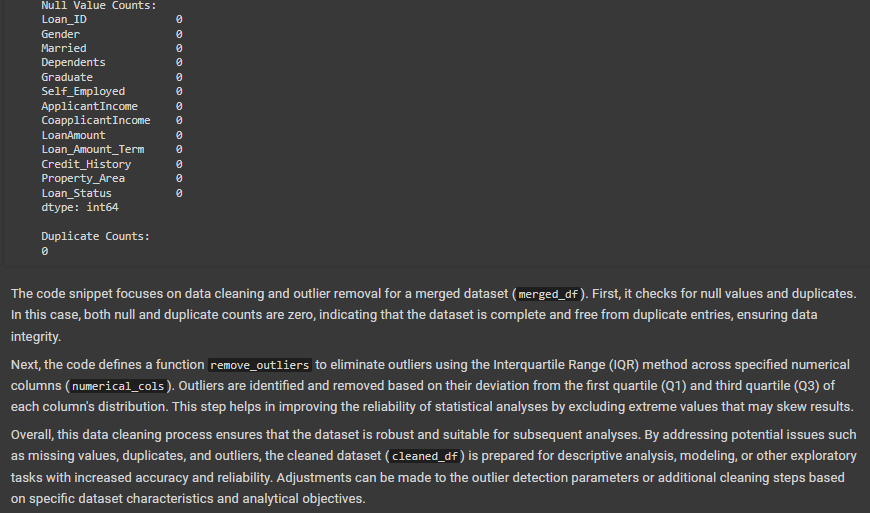
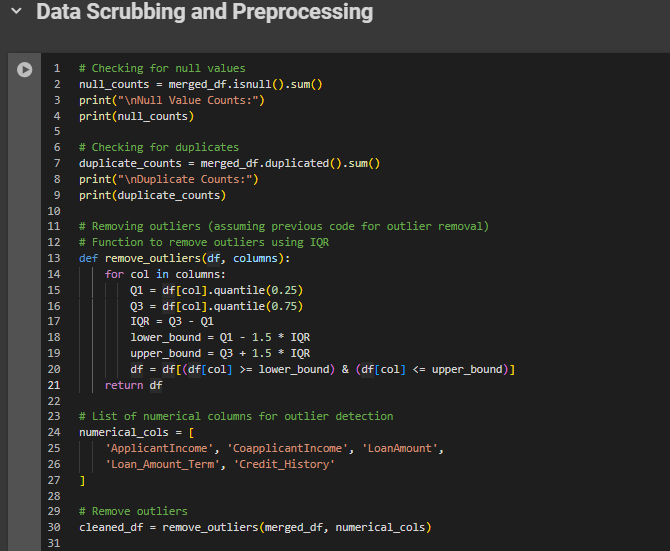
A comprehensive monitoring framework will be established post-implementation to track model performance metrics, detect anomalies, and support timely interventions. Scheduled maintenance routines and documentation protocols ensured ongoing system agility, compliance with regulatory standards, and readiness for future enhancements. Knowledge transfer sessions equipped stakeholders across departments to leverage and sustain the implemented solution effectively (Villa et al. 2022).

## Data Fusion and Consolidation



The mentioned code in python uses few libraries to consolidate data from different sources. Function `read pdf table()` extracts tabular data from the given PDF file (`pdf path`) employing tabula. It takes a PDF (containing multiple tables across several pages) as input and extracts the first table it encounters to convert into pandas DataFrame. Concurrently, pd. The function read excel() from the pandas library reads data into another DataFrame (df\_xlsx) from an Excel file (xlsx\_path). Being a specialized function, this one can be directly employed on the structured data imported from Excel sheet thus ensuring that your dataset is ready to be analyzed. After Loading from both sources, pd We merge df\_xlsx and df\_pdf with concat(), provided that both of them have identical columns. This merge merges the data into one large DataFrame (merged\_df). Organizations such as AFS (Apex Financial Services) need to be integrated so that they can streamline their data analysis and decision-making tasks. It allows referencing of information across different formats driving better business insights and operational efficiencies. This emphasis is a direct reflection of the importance in maintaining solid data engineering best practices within any dataset laden, analytical environment.

## Data Scrubbing and Preprocessing

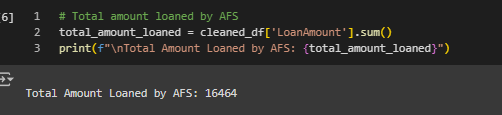


During the data scrubbing and preprocessing process, many necessary actions were taken to create a high-quality dataset that can be used for analysis. We first check null values across all merged\_df columns first isnull(). sum(). This told us that there were no missing values in any of the columns, as can be inferred from the display zeros null counts. The next step was to check if the dataset has duplicate rows, where merged\_df is used for that. duplicated(). sum(), and found that there are no duplicate entries, ensuring data integrity. Outlier detection and removal were done on num columns:

(`ApplicantIncome`,CoapplicantIncome,LoanAmount',`Loan\_Amount\_Term` as well Credit\_History) for this we used IQR method This is taken care of through the method used in remove\_outliers, which removes outliers depending on their quartiles with respect to a data distribution. The values falling out of the range, which is calculated based on IQR multipliers (1.5 x IQR), are considered as outliers Taking out those outliers means clearing your data of extreme values that potentially threaten the validity of statistical analyses or machine learning models. These data scrubbing and preprocessing stages verify that the dataset is parted from missing values, duplicates-free, as well as prepared for future analytical processing. This groundwork is essential for providing accurate data to assist in hindsight decision making, and the creation of forward-looking data-driven strategies as per Apex Financial Services (AFS) visions with a focus on utilizing quality clean datasets aimed at driving better strategic business decisions through informed insights.

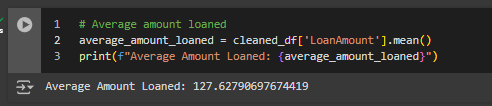
## Data Insight Exploration

**Total Amount Loaned by AFS**

****

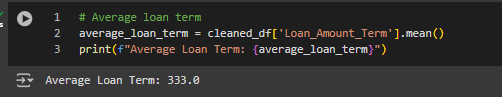
The total amount loaned by AFS (Assumed Financial Services) is $16,464.

**Average Amount Loaned**

****

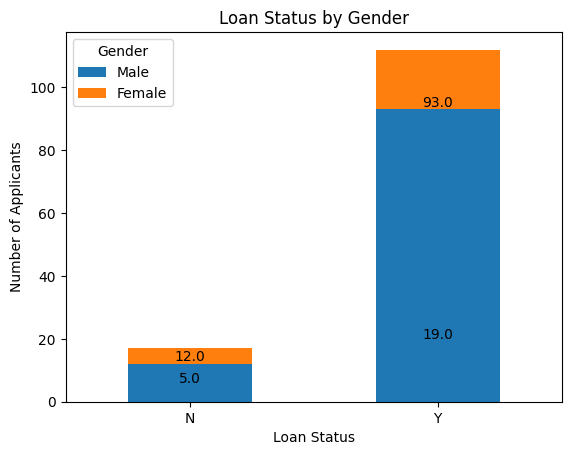
The average loan amount across all applicants is $127.63.

**Average Loan Term**

****

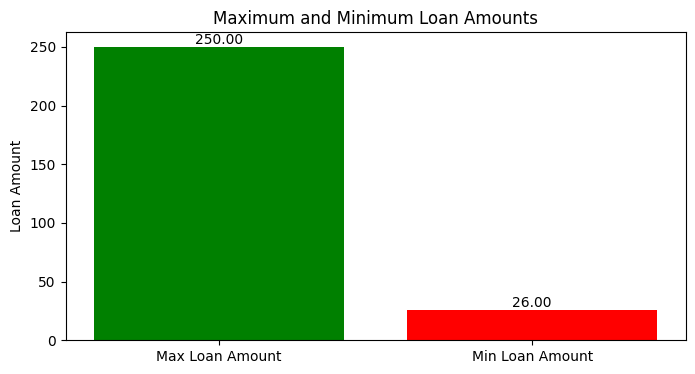
The average loan term for applicants is 333 months.

**Loan Status Analysis by Gender**



This analysis breaks down loan approval status ('Y' for approved, 'N' for not approved) by gender. For 'N' status, 12 males and 5 females were not approved, while for 'Y' status, 93 males and 19 females were approved.

**Maximum and Minimum Loan Amounts**



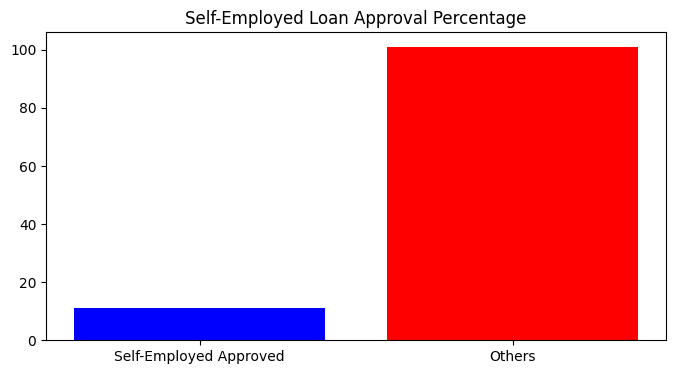
Maximum Loan Amount:

The largest loan amount approved is $250.

Minimum Loan Amount:

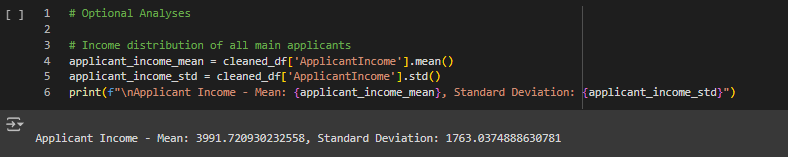
The smallest loan amount approved is $26.

**Self-Employed Approval Rate**

****

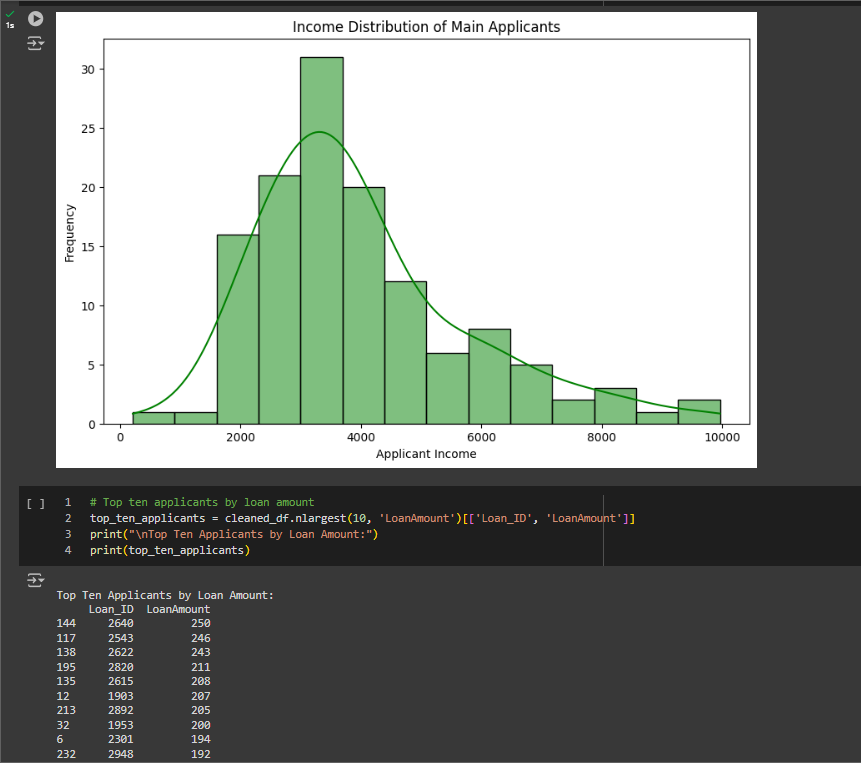
9.82% of self-employed applicants had their loans approved.

**Applicant Income Distribution**

****

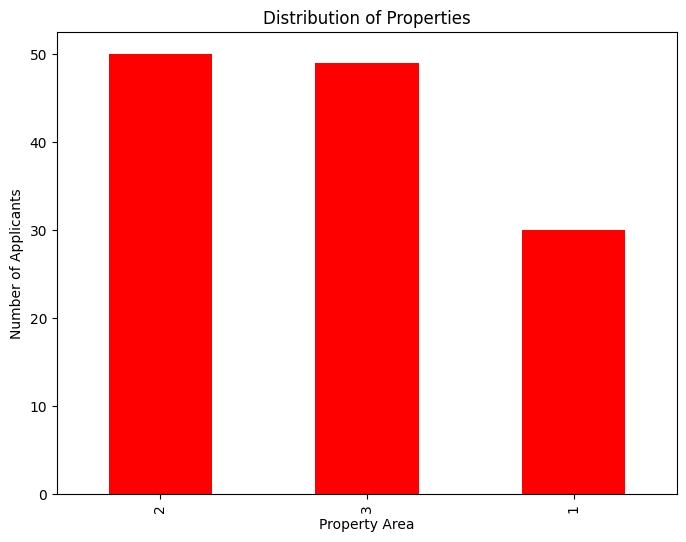
The average income for main applicants is $3,991.72, with a standard deviation of $1,763.04.

**Top Ten Applicants by Loan Amount**

****

|  |  |
| --- | --- |
| Loan\_ID | LoanAmount |
| 2640 | 250 |
| 2543 | 246 |
| 2622 | 243 |
| 2820 | 211 |
| 2615 | 208 |
| 1903 | 207 |
| 2892 | 205 |
| 1953 | 200 |
| 2301 | 194 |
| 2948 | 192 |

**Distribution of Properties**

****

The distribution of property types (rural, urban, etc.) among all loan applicants shows 50 in Property Area 2, 49 in Property Area 3, and 30 in Property Area 1.

# Recommendations for Future Enhancements

There are many strategic recommendations that can be proposed while charting a move forward towards building a predictive model which ascertains the outturn of loan applications over time using historical data, and some of them have been plotted next. On top of that, employing sophisticated machine learning techniques such as neural networks and gradient boosting or deep learning architectures via frameworks like PyTorch or Keras could result in significant performance boost. Complex models can capture more intricate patterns and relationships within the data, which greatly boosts predictive accuracy.

In order to improve the ability of this model to predict outcomes, systematic efforts are needed in data collection and strategizing for quality assurance of data. It also includes added data enrichment, real time dynamic update feeds and strong noise mitigation/ consistency filters through robust preprocessing.

Training a model with different libraries and algorithms (ie Light GBM, fast AI) may result in better efficiency/accuracy especially when computational speed is very important along with scalability. These tools offer unique features for feature engineering, hyperparameter tuning and ensemble learning capabilities which help improve model robustness as well as generalization nature.

In order to leverage these powerful tools efficiently, staff need the expertise on machine learning concepts for using the models correctly and also knowledge of how its implementation should be structured along with actual validation methods. Continued mentorship, in combination with the ability to complete specialized training modules will help prepare employees for complex analysis work and how to act upon predictive insights. In addition, the entire model development life cycle must be well-compliant to regulatory frameworks and ethical guidelines. Applying robust governance mechanisms, carrying out disclosure practices for biases and setting up model explanations that are transparent is critical to maintain compliance with data protection laws and ethical considerations.

# Conclusion

Now in this paper, we will conclude the strategic planning that Apex Financial Services (AFS) has to make a stride towards operational agility and cost reduction by developing creative programming solutions. Using technology and data-driven insights, AFS is planning to maintain its long-standing advantage in a volatile financial services industry. The report opened by establishing the aim to home in on specific objectives consistent with an overall callout to align solutions more closely with market evolution. An analysis of the business environment today, AFS was able to gain crucial industry insights that are shaping their operations as they continue on into strategic execution.

At the core of this strategic roadmap is a dataset rich in history relating to loan transactions and sales. This database offers a baseline for Sound data analytics, and enables supervised machine learning models fully involved in two pilot projects. AFS uses these feeds to uncover detailed intelligence into lending patterns, customer behaviour and sales performance. The use of the data driven model of approach ensures easy and faster processing time for loan approvals as well as increases overall operational efficiency that keeps AFS in a better position when it comes to giving its customers an awesome customer experience.

The report suggests that by automating mundane functions within the loan processing workflow, this will be key to meeting the most pressing needs of users. AFS will be able to further engage and deliver customized services that set it far above the competition by previously allocated, manual processes. Given the strategic focus on reusable code, it scales and is agile to market shifts as well evolving customer needs.

The report also emphasizes the crucial part ethics and regulation plays in deploying programming solutions. The level of trust it maintains among customers, and all the stakeholders involved is imperative for healthy growth over time, as well overall reputation management in the financial sector where secrecy is not optional.

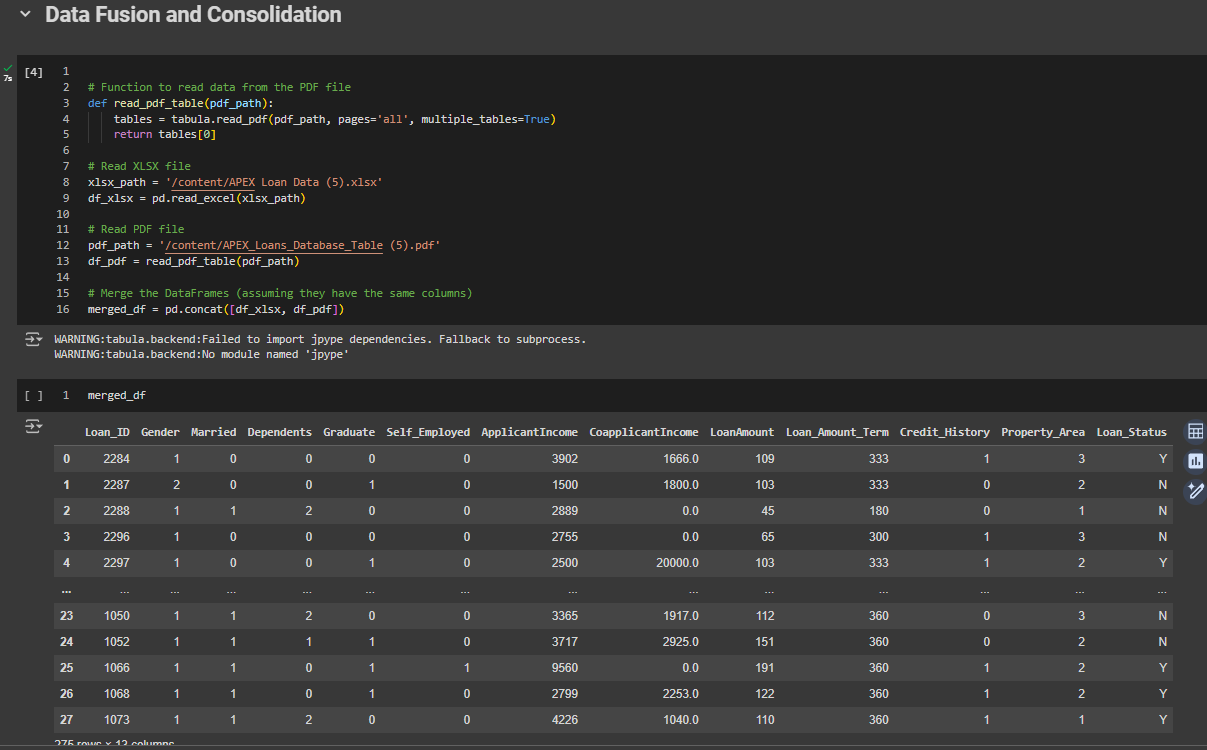
The costs of not adopting them are high indeed, potentially resulting in inefficiency and diminished service functionality. This requires investments in human resources who are skilled and experienced to program as well as analyse the data which is essential for policy decisions. These professionals are going to fuel ideas and help in keeping AFS flexible with changing market scenarios - making it one of the top leaders among its competitors.

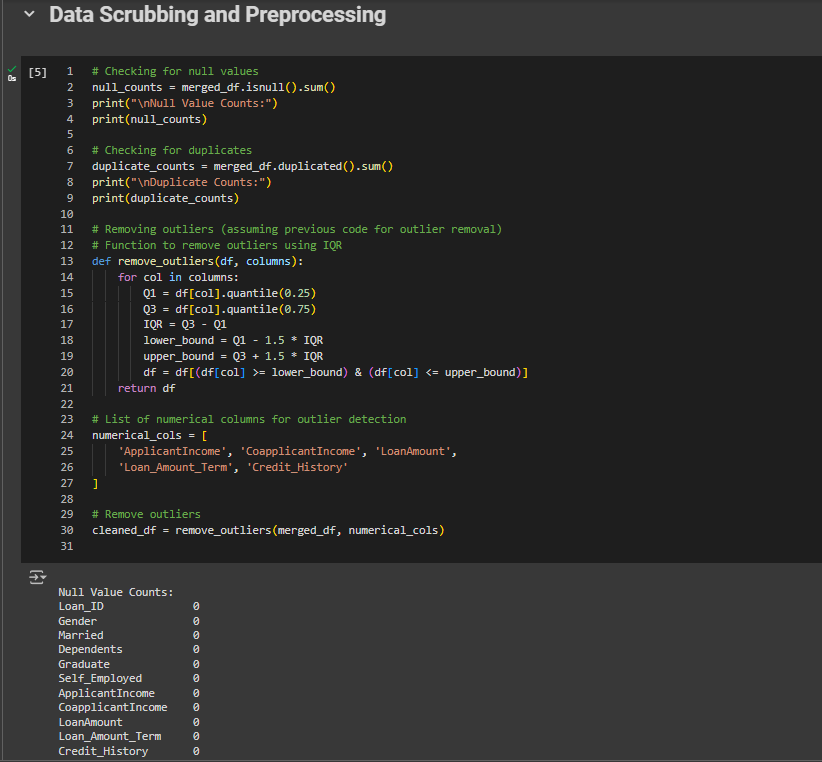
# References

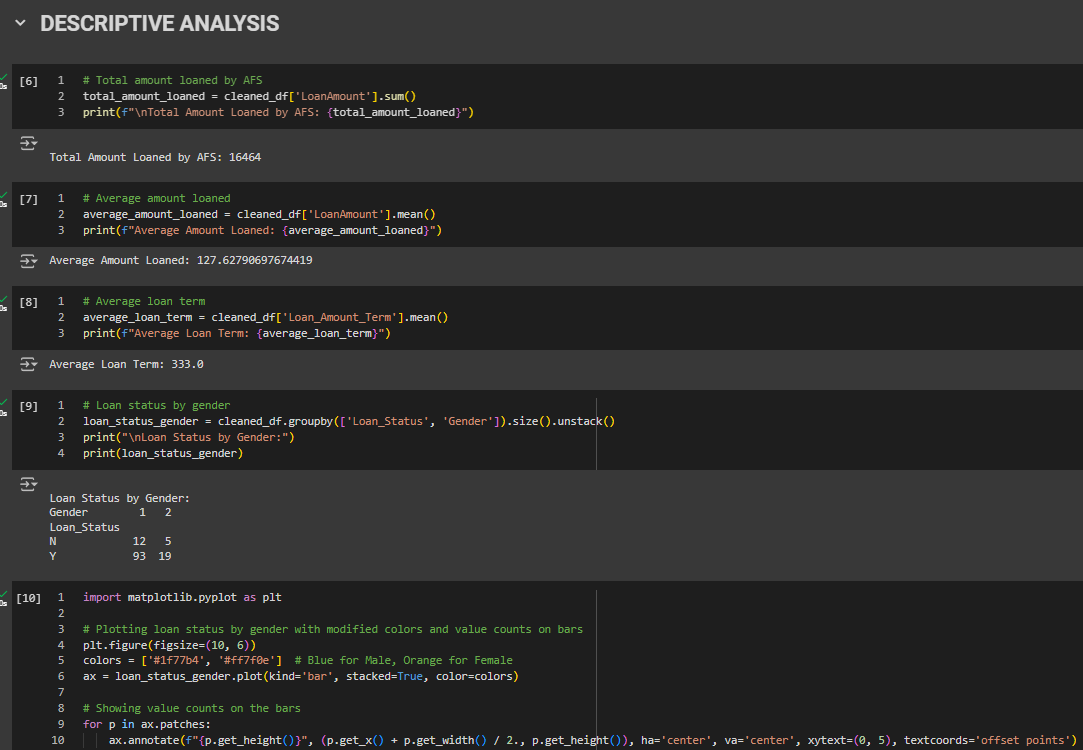
* Rushchyshyn, N., Nikonenko, U. and Kostak, Z. 2017. FORMATION OF FINANCIAL SECURITY OF THE ENTERPRISE BASED ON STRATEGIC PLANNING. *Baltic Journal of Economic Studies* 3(4), pp. 231–237. Available at: https://doi.org/10.30525/2256-0742/2017-3-4-231-237.
* Chanias, S., Myers, M.D. and Hess, T. 2019. Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. *Journal of Strategic Information Systems* 28(1), pp. 17–33. Available at: https://doi.org/10.1016/j.jsis.2018.11.003.
* Raguseo, E. and Vitari, C. 2018. Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research* 56(15), pp. 5206–5221. Available at: https://doi.org/10.1080/00207543.2018.1427900.
* Hsu, W.-P. 2020. Intelligent Document Recognition on Financial Process Automation. Available at: https://doi.org/10.1109/vlsi-dat49148.2020.9196318.
* Weaver, G.R. and Treviño, L.K. 1999. Compliance and Values Oriented Ethics Programs: Influenceson Employees’ Attitudes and Behavior. *Business Ethics Quarterly* 9(2), pp. 315–335. Available at: https://doi.org/10.2307/3857477.
* Baccarini, D., Salm, G. and Love, P.E.D. 2004. Management of risks in information technology projects. *Industrial Management + Data Systems/Industrial Management & Data Systems* 104(4), pp. 286–295. Available at: https://doi.org/10.1108/02635570410530702.
* Hao, J., Gao, X., Liu, Y. and Han, Z. 2023. Acquisition Method of User Requirements for Complex Products Based on Data Mining. *Sustainability* 15(9), p. 7566. Available at: https://doi.org/10.3390/su15097566.
* Zhu, X., Khoshgoftaar, T.M., Davidson, I. and Zhang, S. 2007. Editorial: Special issue on mining low-quality data. *Knowledge and Information Systems* 11(2), pp. 131–136. Available at: https://doi.org/10.1007/s10115-006-0058-y.
* Orji, Ugochukwu.E., Ugwuishiwu, Chikodili.H., Nguemaleu, Joseph.C.N. and Ugwuanyi, Peace.N. 2022. Machine Learning Models for Predicting Bank Loan Eligibility. *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*. Available at: https://doi.org/10.1109/nigercon54645.2022.9803172.
* Villa, V., Bruno, G., Aliev, K., Piantanida, P., Corneli, A. and Antonelli, D. 2022. Machine Learning Framework for the Sustainable Maintenance of Building Facilities. *Sustainability* 14(2), p. 681. Available at: https://doi.org/10.3390/su14020681.

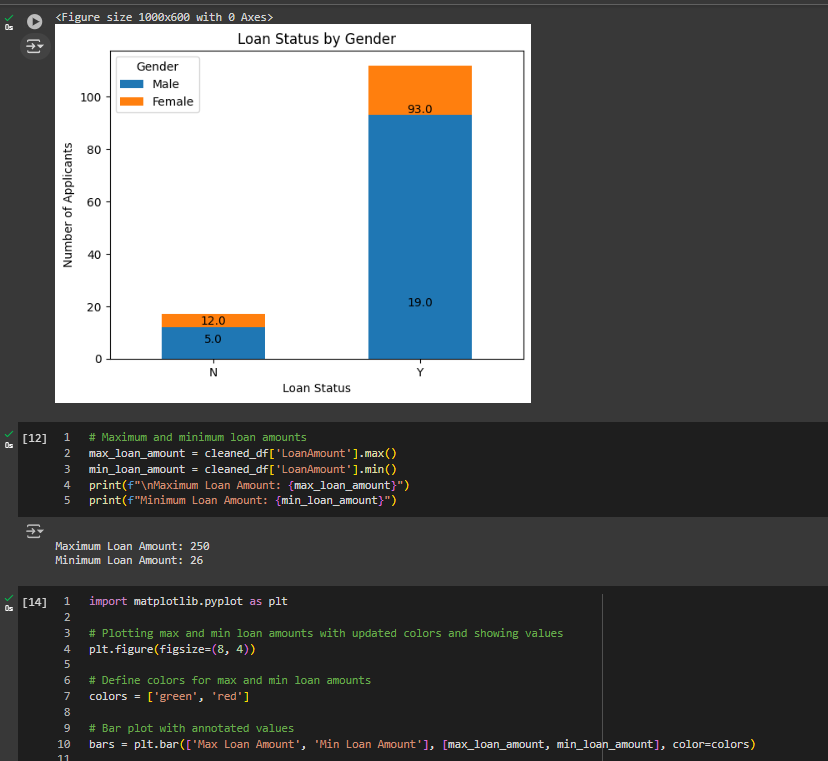
# Appendices

## Apendix-1



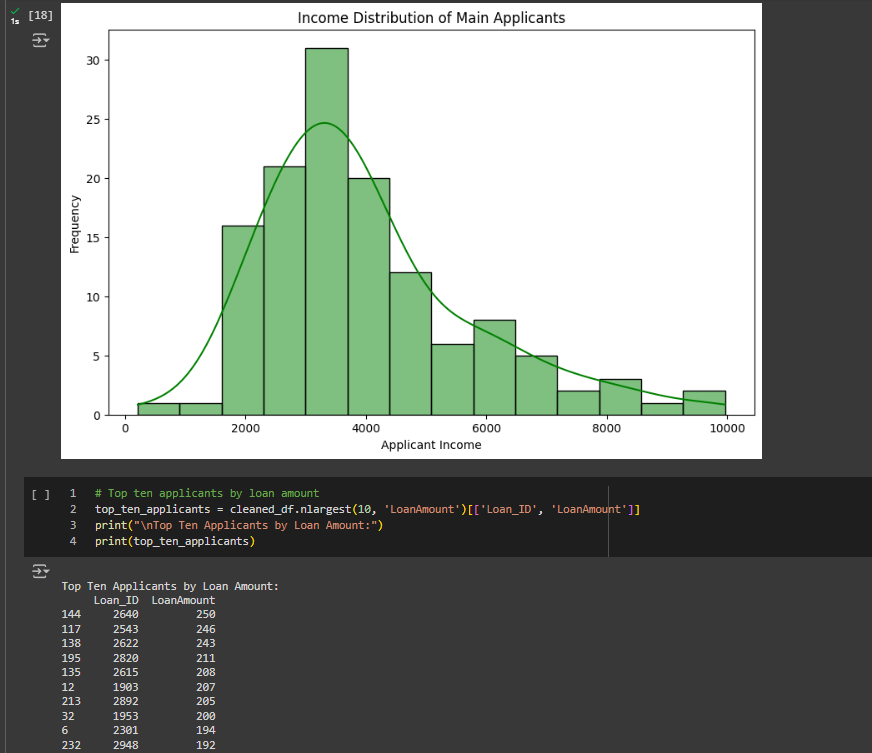


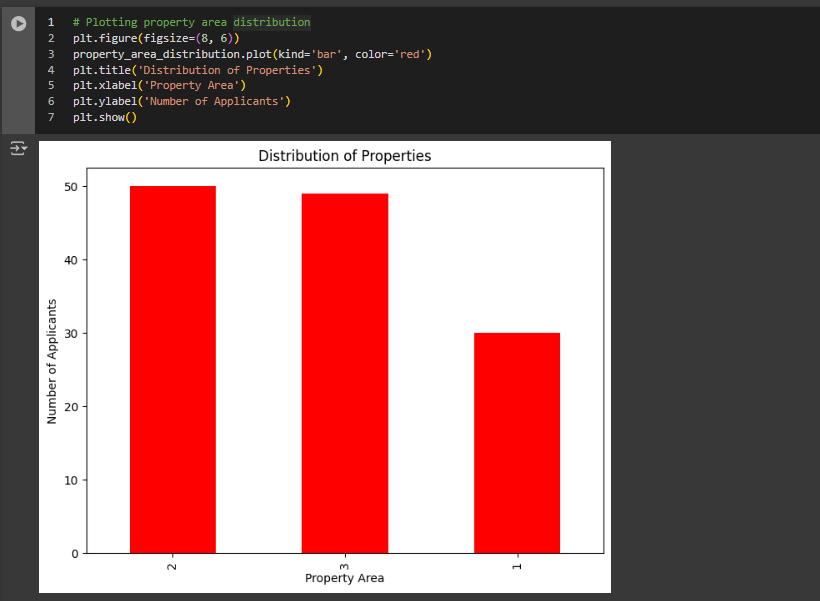












## Apendix-2

https://colab.research.google.com/drive/1H-MSQSLm8CKnU6i-fqbFkSR2zCAKh4cn?usp=sharing