Peer-to-Peer Lending: Investor Dashboard

Team 14: Clayton Francis, Nicholas Park, Robson De Souza, Xiaolu Su, Bingqian Hu

1 INTRODUCTION

The Peer to Peer (P2P) lending market is a decentralized platform where individuals borrow and lend with one another. Founded in 2006, Lending Club was one of the first P2P platforms to publish an API to market researchers. Our goal is to use historical loan data to model the probability of default and create recommended diversified portfolios with interactive credit risk metrics for a set of test loans. Currently, there is some academic research on P2P lending portfolio diversification. Still relatively new, P2P lenders are often under-informed, investing in risky uncollateralized loans. Quality independent research is not enough, instead lenders need tools to guide well-informed investment decisions.

Our approach is to introduce a platform for lenders to interact with a portfolio modeling assistant. Recent increased activity in decentralized finance and retail investing suggests new tools will help attract lenders to the P2P community. Investors can better assess risk, platforms can market increased transparency, and more loans will be funded, resulting in a social impact on borrowers. Success is measured by platform traffic, user feedback on technology, and performance of suggested portfolios.

Potential risks include discontinued API, scalability, and poor model performance. Payoffs include investment efficiency, expanding P2P community, and potential platform integration. The prototype dashboard will have 0 cost, but future implementations may. Building the proof of concept will take approx. 2.5 months. The midterm checkpoint is a simple functioning dashboard whereas the final exam is the ability for new users to install the platform and use the visuals to interact with our models.

2 PROBLEM

When a bank denies someone a loan, where can they turn next? Family, friends, community? P2P lending opens doors for those who otherwise would not receive funding for personal expenditures. On the other hand, investors are always searching for alternative ways to allocate capital and diversify their portfolio. Rather than invest in corporate or government bonds, investors can abundantly diversify their portfolio by lending to individuals in communities across the world.

So, what's the problem? Well, the P2P lending market is still a relatively under-invested market given it's potential size. Those who do invest often fail to effectively diversify because they are either emotionally funding loans based on qualitative factors, or over-pursuing excessively high returns which can often lead to default.

We believe a key to increasing funding on P2P platforms is to further educate investors. Our proposed innovation is to build an interactive dashboard where investors are informed of overall market analytics and assisted by tools to help them build optimized diverse portfolios. Continued growth in P2P markets would have substantial positive social impact where both borrowers and lenders could simultaneously benefit.

3 LITERATURE SURVEY

Restrictive monetary policy yields investors a higher return for less risk, making it difficult for risky borrowers to attract funding. Our project seeks to harness innovations in lending against current economic conditions. Pure research is not enough as we look to offer a product that encourages lending [15]. To minimize risk for a given level of expected return, investors diversify their portfolios. Our project notes the unsecuritized nature of P2P lending increases risk, and thus the challenge is to quantify risk via loan outcome probability [7]. Sharpe ratio can be used to maximize return for a given level of risk, historically for equity funds. We will evaluate our sample portfolios with Sharpe Ratio using expected returns and volatility rather than the typical calculation which looks at historical time series data [4].

Emotion triumphs data for many investors. Our project believes lenders should rely solely on modeled historical performance of loan attributes. This paper needs to consider solutions, instead of just highlighting the issue [8]. To account for emotional bias, P2P lending site Prosper made platform changes to reduce qualitative loan information which successfully increased platform volume. This paper reinforces our proposal as it highlights the success of focusing on core loan characteristics. We plan to build on this by proposing ways to reduce emotional bias [5]. Another paper investigated which loan features drive P2P loan prepayment. This research will help us determine significance in our models, and we can take it a step further by letting users interact with the results [12].

One paper discusses random forest classification methods for predicting borrower status and compares the performance against support vector machine (SVM) and logistic regression. Machine learning is widely used by financial institutions to predict risks and payoffs of potential investments. SVM is not suitable for large data sets and logistic regression achieves lower performance than peer methods [9]. Artificial neural networks (ANN) are used to create a credit scoring model in classifying P2P loan applications into categories of default. However, ANN predictions lack transparency, which makes it difficult to explain the results [1]. ANN tends to perform highly when applied to unstructured data, but for tabular data, decision tree algorithms are currently considered best-in-class. Upon analyzing many ML methods, we look to implement the modern LightGBM and XGboost algorithms. We didn't identify any potential shortcomings since these are the preferred models we will adopt in this project [10].

Various ML estimators are recommended to model credit default probability. Our project will experiment with the performance indicators discussed in this paper. Vanilla models are not always performant, so our proposal will experiment with a stacked ensemble of classifiers to improve AUC. [11]. Model understanding suffers as complexity increases, and this paper reviews ways to visualize and explain the various effects that features have on predictions. Our project will use the suggested SHAP visualization to explain the model predictions. The paper does not delve into the notion of AI fairness, which we will look to address. [2]. The next paper extensively reviews the methods of pricing loans using default probability. This is helpful background knowledge for our project. However, the paper fails to capture this process into a fool-proof loan grading scheme like Lendingclub does, which we will use to compare against own model predictions [13].

Javascript's D3 is a powerful visualization library when combined with Python. D3's ability to customize visuals could help better-explain the complicated outputs of our risk models. However, implementing customization requires a level of expertise that would drastically lengthen our timeline, compared to other easy-to-use software [3]. Power BI has both the ability to integrate with Python and quickly publish interactive visualizations. This is highly attractive to our short time-lined implementation. One limitation is that Python visualization is not supported for web publishing [14]. Tableau efficiently creates visualizations and graphics from large datasets used for machine learning and big data applications. The last paper uses Tableau to highlight the benefits of recommender systems in visualization tools. We argue such tools promote a lesser understanding of the data and put the analysis at risk. Although we will use similar software for visuals, we will not rely on recommender systems, rather spend the quality time becoming experts of our data [6].

4 PROPOSED METHODS

4.1 Expected Innovation

As highlighted in the previous sections, investors need innovative tools to help safely guide them into this new peer lending market. Researchers agree that diversification and social impact makes P2P lending an exciting investment opportunity, but how does one begin? The current state of the art P2P platforms obviously allow for investors to diversify, but designing a basket of loans is entirely subjective to the user.

Our objective is to use Lending Club data to train a model predicting default probability. Lending Club and similar market participants develop proprietary models for determining credit risk, assigning a rating to each loan (A, B, C, etc.). We look to build on this concept by observing the true outcome of past loans in each category to ultimately model future performance.

Upon forecasting default probability we use a set of test loans to build optimized portfolios based on our own model predictions. We create multiple sample portfolios for different levels of risk and diversification selected by users in an interactive PowerBI dashboard. The dashboard also showcases a market overview, risk metrics, and feature importance; allowing investors to easily choose data-driven options, rather than relying on emotional intuition or guessing.

4.2 Approach Description

4.2.1 Credit Risk Model. As described in the previous section, P2P lending platforms typically have proprietary risk models which they use to grade newly issued loans. The loan grading and interest rate gives investors a relative idea on the riskiness of the loan, but we take this a step further by observing the actual outcomes of historical P2P loans and providing a deeper understanding of the driving factors of default. The output of our credit risk model is probability of default, and the input consists of the borrower's credit bureau variables and demographic features that contribute to his or her creditworthiness.

Here are a few assumptions we inject into the data. We limit the scope of the dataset by filtering for loans issued after 2015, so our model is trained on the recent lending landscape. Also, we limited the loan terms in scope to 36 months (3 years). The so-called "current" loans without a final determination (i.e., neither fully paid nor defaulted) are also excluded. Lastly, further data cleaning was performed to remove "look-ahead" features that are not available to us at the time of underwriting (e.g., recovery amount) as well as features that were unequivocally insignificant in predicting default (e.g., "url"). This data is then split: 1) the most recent 10,000 loans reserved as test set for portfolio building, 2) the next most recent 10,000 as our validation set and 3) all other loans as our training set.

Following data preparation, we began fitting the data pipeline that can be used to real-time preprocess and predict new incoming data. A class was built to instantiate the full-pledged data pipeline, harboring the best ML algorithm along with many utility functions for visualization, validation and analysis. A call to the constructor uses a .yaml config file to specify hyperparameters and other configurable rules. The model pipeline using imblearn/sklearn entails 1) feature preprocessing (e.g., imputing, scaling, string-formatting), 2) under/oversampling to handle heavy class imbalance, 3) lasso feature selection and 4) ML model fitting - which are all collectively fitted to the train data in our custom "fit" method within the class. Jupyter notebooks were used to call the class and utility methods to train multiple models and evaluate which model yielded the highest accuracy scores. For model experimentation, leveraging the autoML package PyCaret helped us identify that CatBoost and other gradient boosting models outperforms peer classifier models such as neural networks. Finally, we implemented a stack ensemble of CatBoost models along with a grid search of hyperparameters for marginal AUC improvement.

After validating the model on observed loan performance, we also evaluated each independent variable's impact using the SHAP (SHapley Additive exPlanations) library, a visualization based on game theory to help explain ML model outputs. The next sections dive deeper into how the model's predictions on the test set are used to build optimized portfolios and is showcased to investors in an interactive tool.

4.2.2 Portfolio Assistant. Although building our own credit risk model is an impressive innovation, it more so impacts the "behind-the-scenes" analysis. Our innovative product offering directly used by investors is an interactive dashboard with a portfolio building assistant designed to reduce emotional investing and over-funding risky loans for the uninformed hope of obtaining higher returns.

Our goal was to make building a diversified portfolio as easy as possible for investors, and provide suggestions for optimized portfolios based on investor preferences. Once a potential investor pulls up the dashboard, they can select from a list of preset values including the number of loans to diversify across and choosing from a combination of Lending Club loan grades to define their risk appetite. The optimized portfolio tool then displays the list of suggested loans to fund, the expected return, cash flows, Sharpe ratio, and resulting level of diversification.

To create the preset optimized portfolios we imported the set of test loans into a Jupyter notebook which built an optimization problem for each combination of user selections. Using the "pulp" package we could define constraints for the loan grades and number of loans. The target for the optimization problem is to maximize the performance of a portfolio by selecting loans with the highest expected return where each portfolio is limited to a unique set of constraints. The formula we define for expected return is represented as...

$$(1 + i[x]) * (1 - DP[x])$$

where i = interest rate, DP = default probability (predicted by model), and x represents each loan in the portfolio. This objective function rewards loans with the highest return, discounted by the default probability predicted by our credit risk model. After setting up the constraints and objective function, the pulp package implementation chooses the optimal loans for each portfolio. A table of resulting portfolios is exported to Excel and loaded into the dashboard.

As a result, investors have immediate access to a simplified selection of optimized diversified portfolios driven by an ensemble of high performing boosted machine learning credit risk models trained on the historical performance of similar characteristic loans. Doing the work for investors and presenting them with preset options helps reduce uncertainty and emotional investing, and educates about the power of diversification. Growth in P2P lending is dependent on widespread acceptance driven by better informed lenders gaining comfort-ability with lending to peers and being rewarded with positive returns. (Appendix A) It should be noted that investors have the ability to partially fund loans with minimum amounts of $25, giving investors at all levels access to the suggested portfolios and potential return.

4.2.3 Market Insights. The third innovation is a separate part of the PowerBI dashboard which provides an overview of the P2P market. The Market Insights tab in the dashboard imports the post-processed model output and shows high-level summary statistics of loan and borrower characteristics, feature importance visuals, and trend charts of how loan performance differs across key sub groups of model inputs.

Since P2P borrowers do not post collateral, the chances of a lender losing their entire investment is much greater than other fixed income products. To foster recurrence, investors must be better informed upon entering this new market. Consolidating statistics on current market structure and using data-driven evidence to explain which features are critical to loan performance is an overlooked but extremely beneficial innovation which needs to be reiterated on all platforms. (Appendix B)

5 EXPERIMENT EVALUATION

5.1 Credit Risk Model Experiment

The final version of our credit risk model was selected after we tested many different ML classification algorithms and combinations of hyper-parameters to minimize the AUC error when making predictions on the testing and validation data sets. Following this process, we technically choose the "best" performing model out of the ones we implemented, but does this actually mean the model is a high performing predictor? How does our model's predictions compare to Lending Club's proprietary loan grading model?

Essentially, to confirm that our model was fitting properly to the overall problem, we needed to conduct an experiment to somehow compare our model predictions to Lending Club's view on the loans. To do this, we made the important distinction that our probability of default output values (in a range of 0 to 1) are a way to measure risk. Lending Club's determination of risk is represented by the interest rate applied to the loan after evaluating the borrower features. Both default probability and interest rate are ways to assign risk to loans, so by comparing the values of our definition of risk to Lending Club's, we can measure the overall effectiveness of our model.

To compare the two different measures of risk, we calculated the correlation between our predictions of default probability and the loan interest rates for the test data. The correlation coefficient was 0.43 (positive with moderate strength), which proves that our model outputs reasonable predictions. We interpret that the remaining differences can be explained by the additional features factored into our model's attempt at making predictions on the true outcome of the loans, along with a different algorithm implementation than Lending Club. Although the correlation is positive and significant, it is always a nice check to view the results in a graph. Since there are 10,000 loans, plotting all of them would display too much noise to interpret. Instead we binned the loans into groups based on different ranges of default probability, and by taking the average interest rate of each group, we get a correlation plot that better visualizes the relationship between the two risk metrics, and supports the conclusion that our model is relevant and effective. (Appendix C)

This experiment confirms the model is trained properly and is not somehow making unrealistic predictions that happen to yield the lowest error. Additional performance testing was performed on the validation set to further confirm our model fits to unseen data. As previously mentioned, we also use the SHAP visualization to show the most significant model features such as # of mortgage accounts, # of personal finance inquiries, and # of finance trades. (Appendix D)

5.2 Optimized Portfolio Experiment

Next, using the predictions from our credit risk model, we were able to build a preset list of optimized portfolios which investors can select depending on their preference for risk and level of diversification. We needed a way to validate that our suggested portfolios are successful, to gain confidence that they will outperform an average investor. So how do we know that our "optimized" portfolios actually perform well? Can we define a benchmark for comparison? Do the optimized portfolios beat the benchmark?

To answer these questions and essentially determine that our list of suggested optimized portfolios are "better" investments, we need to calculate their actual performance and compare it to some benchmark. For each portfolio, we built a benchmark portfolio consisting of 5 random loans all having a similar risk profile. Similar to the construction of the optimized portfolios, the benchmark portfolios can only select loans from the test data set (most recent 10,000 loans), and are constrained to loans with an interest rate +/- 3 percent the average interest rate of the comparable optimized portfolio.

We designed the random portfolios with the goal of replicating the loan selection of an uninformed investor. In theory, such individuals will have some level of understanding about how much risk they are willing to take, but cannot scientifically account for all the features that our credit risk model does, and essentially they choose the individual loans at random. The random portfolios only contain 5 loans because P2P investors typically don't take full advantage of the vast opportunity for diversification that P2P lending offers. For comparison, the optimized portfolios consist of either 5, 10, 15, or 20 loans. Our hypothesis is that on average the optimized, diversified portfolios will outperform the randomly chosen less diversified portfolios of a similar risk profile.

Table 1 shows the key performance metrics we used to compare the portfolios, resulting in the optimized portfolio being the clear winner as it outperformed the random portfolio in every category. On average, the optimized portfolio yielded much higher returns (random portfolio returns were negative), much less underlying loan defaults, and less volatility of returns. It should be noted that each run of the optimizer can yield different portfolios resulting in different outcomes and statistics. If the jupyter notebook is re-run the results may not match the ones shown above or in the latest version of the dashboard. We confirmed after various optimization runs that the outcomes and conclusions discussed in this paper mimic the typical results.

Metric	Optimized	Random
Avg Return	3.75%	-3.07%
Avg Months to Paid/Default	2.54	2.54
Num of Defaulted Loans	0	18
Avg Annualized Return	18.35%	-7.45%
Best Portfolio (Annualized)	38.0%	30.0%
Worst Portfolio (Annualized)	8.0%	-100%
Outperforming Potfolio Count	44	16
Avg Sharpe Ratio	2.70	-0.17

Table 1: Optimized Portfolio Experiment Results

Since the random portfolio only has 5 loans (less diversification) it is subject to a larger standard deviation of returns. Greater volatility typically results in a higher risk of loss, which the random portfolio achieved by having at least 1 portfolio lose the entire investment, compared to the worst optimized portfolio actually still gaining 8%. Sharpe ratio, a common portfolio performance metric accounting for the trade-off between risk and return, shows how overall significantly better the optimized portfolio tends to perform.

Lastly, since each portfolio matures over different time horizons we realized we needed to calculate annualized return for a cleaner comparison. This idea lead us to a very interesting discovery that we had not yet observed in the data. On average, the underlying loans in both the optimized and random portfolios tended to either be fully paid off or go into default within the first few months of issuance. Prior to this experiment we were under the assumption that loans of 36 or 60 month terms were in repayment for the majority of time.

Such information helped draw additional conclusions about borrowers which can be helpful to investors in this space. In general, even though loans are issued with 3 or 5 year maturities, the expected time horizon for most investments can be seen as just a few short months. Typically, a shorter loan maturity is positive for investors since they are then able to collect and have continuous access to their capital, hopefully reinvesting and compounding their return. Additionally, when annualizing the realized returns seen over just a few months, the average return of an optimized P2P portfolio looks to be an extremely attractive investment compared to other fixed income or equity products, building a stronger case for the growth of P2P lending.

6 CONCLUSION

We conclude that the Peer to Peer Lending market is a sufficient alternative investment that deserves greater attention and increased funding. Too many investors currently utilize P2P lending as a way to identify high yield opportunities, rather than perform adequate data modeling or implement diversification.

At first glance, lending to strangers without collateral seems like a frightening concept. We hope that our market research, summary data visuals, default probability model, and interactive portfolio assistant all contribute to furthering the understanding of market participants and fuel the continued growth of P2P lending.

P2P lending sites such as Lending Club provide invaluable data which we believe should drive investor decision making. Many of the average individual investors first entering this market do not have a computer science or machine learning background. Rather than funding loans based on instinct, we believe it is absolutely imperative that all investors have access to validated statistical analysis which accounts for data features that humans cannot.

The successful implementation of our gradient boosted credit risk model is the core component of our project. Our pipeline is designed to process new incoming data and predict a loan's outcome more accurately than any generic risk profile assigned by the lending platforms. Using actual historical P2P loan outcomes to model the true deterministic factors driving loan performance should provide a newfound confidence to all investors, not just those with a background in data modeling.

Existing research typically ends after drawing some initial conclusions, but we did not stop there. Higher usage rate for new applications typically occurs when individuals are spoon-fed information rather than forced to spend time figuring out details. As a result, we built an easy-to-use tool so investors can now quickly gain a deep understanding about key market factors, in addition to portfolio suggestions designed to minimize the probability of default while maximizing return. Although it is suggested that investors perform their own additional due diligence, in theory someone without any prior knowledge of P2P lending could now simply invest in one of our portfolios, and according to our experiments they will more than likely outperform the market.

The team's combination of research, analysis, experimentation, and collaboration helped foster many successful ideas, outcomes, and learning experiences for ourselves and others to build upon in the future.

All team members have contributed a similar amount of effort

REFERENCES

[1] Byanjankar A and Mezei J. 2015. Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. *IEEE Symposium Series on Computational Intelligence* (2015). http://vigir.missouri.edu/~gdesouza/Research/Conference_CDs/IEEE_SSCI_2015/data/7560a719.pdf

[2] Misheva B and Osterrieder J. 2021. Explainable AI in Credit Risk Management. *Risk Management* (2021). https://arxiv.org/abs/2103.00949

[3] Kyran D. 2016. *Data visualization with python and javascript: scrape, clean, explore transform your data* (2nd. ed.). Wiley, New York, NY.

[4] Sharpe W F. 1966. Mutual Fund Performance. *The Journal of Business* 39, 1 (1966), 119–138. https://www.jstor.org/stable/2351741

[5] Ryan S G and Zhu C. 2018. Fintech Isn't So Different From Traditional Banking: Trading Off Aggregation of Soft Information for Transaction Processing Efficiency. *ssrn electronic journal* (April 2018). http://dx.doi.org/10.2139/ssrn.3212902

[6] Kevin H and Michiel B. 2019. Data analytics and visualization using Tableau utilitarian for COVID-19. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (May 2019), 1–12. https://doi.org/10.1145/3290605.3300358

[7] Markowitz H. 1952. Portfolio Selection. *The Journal of Finance* 7, 1 (1952), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

[8] Riggins F J and Weber D M. 2017. Information asymmetries and identification bias in P2P social microlending. Information Technology for Development. *Information Technology for Development* 50, 1 (April 2017), 107–126. https://doi.org/10.1080/02681102.2016.1247345

[9] Malekipirbazari M and Aksakalli V. 2015. Risk assessment in social lending via random forests. *Expert Syst* 42, 10 (June 2015), 4621–4631.

[10] Xiaojun M and Jinglan S. 2018. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications* 31 (2018), 24–39. https://www.sciencedirect.com/science/article/abs/pii/S156742231830070X

[11] Addo P and Guegan D. 2018. Credit risk analysis using machine and Deep Learning Modelss. *Risks* (2018). https://doi.org/10.3390/risks6020038

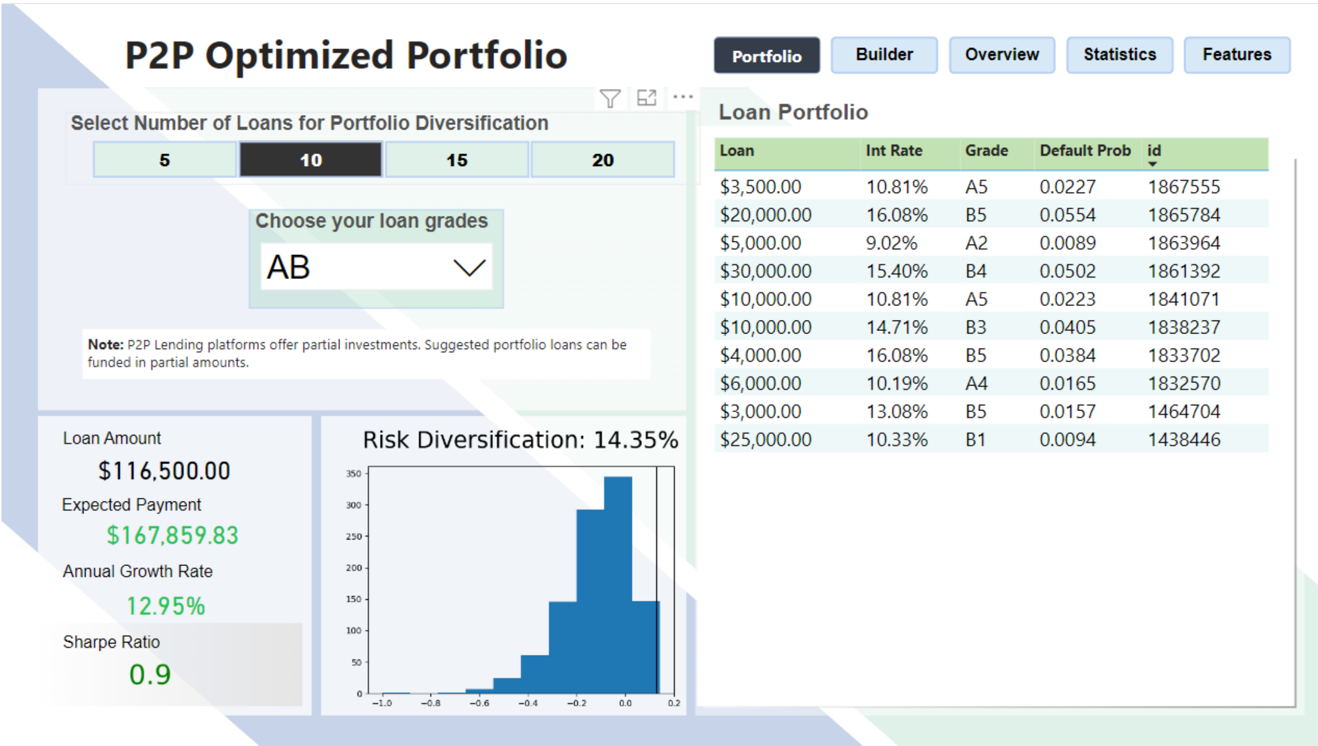
[12] Li S and Li X. 2017. Which Factors Drive the Prepayment of P2P Lending? Evidence from Prosper.Com. *ssrn electronic journal* (2017). http://dx.doi.org/10.2139/ssrn.3054424

[13] S. M Turnbull. 2003. Pricing loans using default probabilities. *Economic Notes* 32, 2 (2003), 197–217. https://doi.org/10.1111/1468-0300.00111

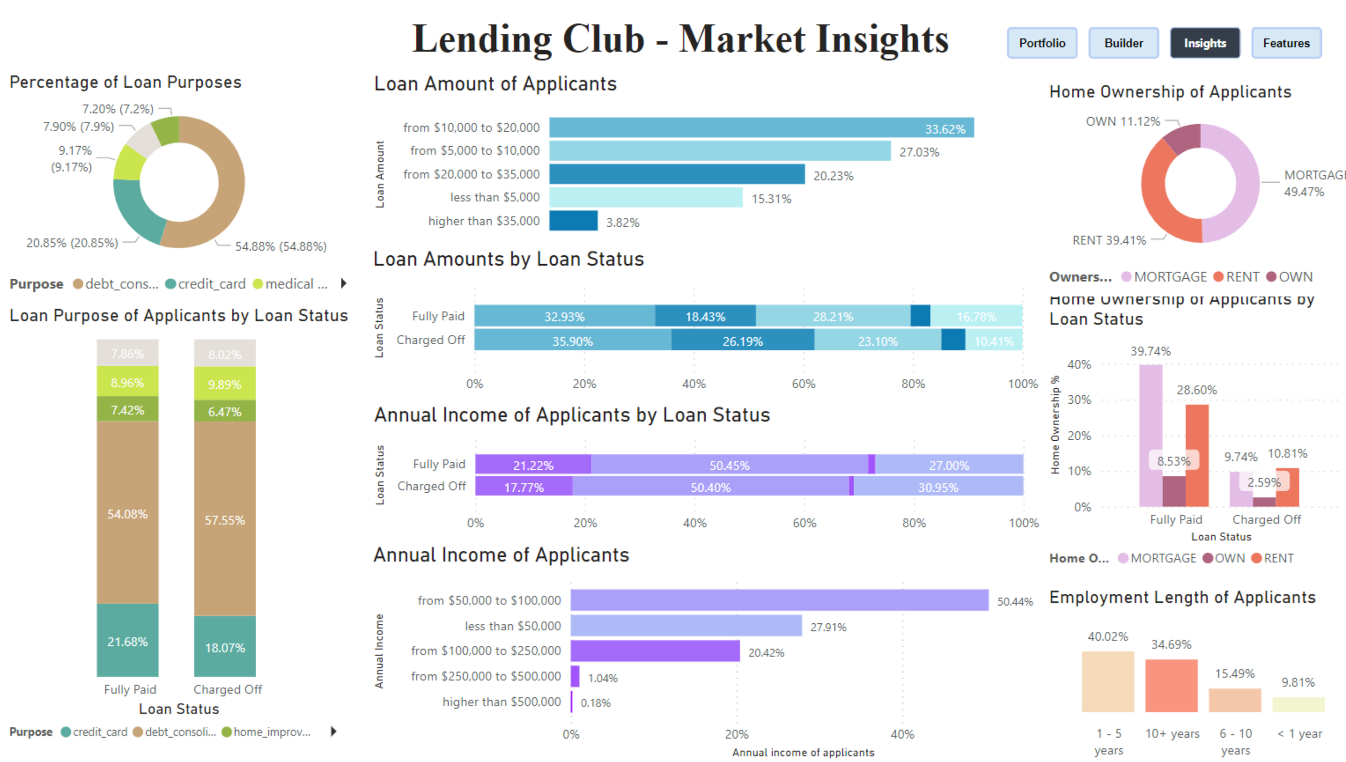
[14] Vashisht V and Jakhmola N. 2021. An Effective Approach for Integrating Microsoft Power BI Application with Python for Predictive Analytics. *Micro-Electronics and Telecommunication Engineering* (2021). https://doi.org/10.1007/978-981-33-4687-1_45

[15] Chu Y and Deng X. 2019. The Risk-Taking Channel of Monetary Policy: Evidence from Individual Investors in the Peer-to-Peer Lending Market. *ssrn* (2019). https://doi.org/10.2139/ssrn.3171676

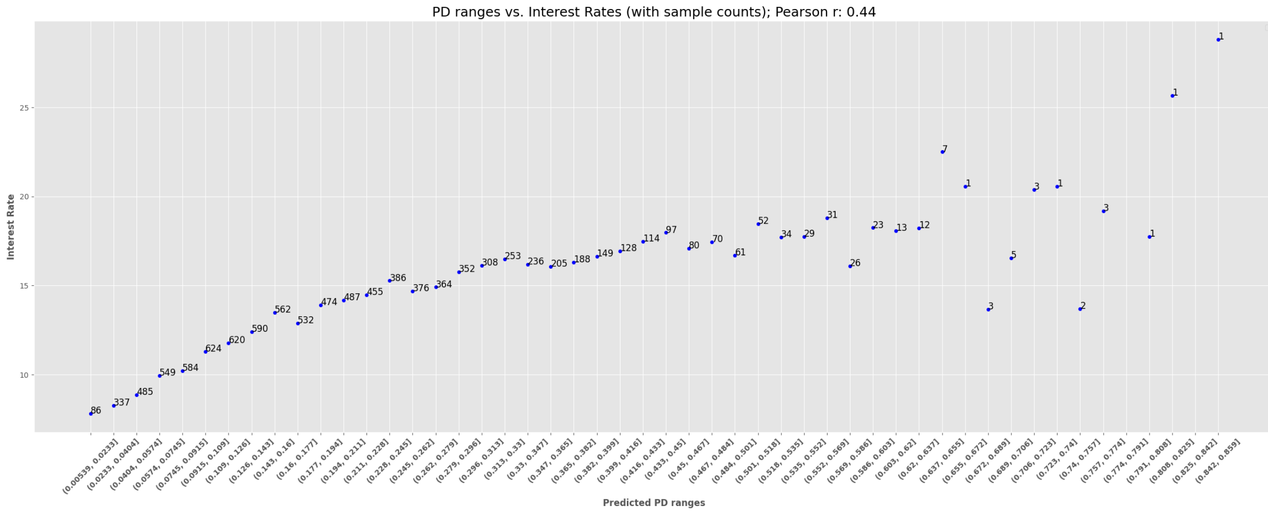
A PORTFOLIO ASSISTANT SCREENSHOT



B MARKET INSIGHTS SCREENSHOT



C DEFAULT PROBABILITY VS INTEREST RATE



D FEATURE IMPORTANCE - SHAP VISUALIZATION