

**INSIDE THE BAZAAR:
UNVEILING THE LIVES OF
LOCAL SHOPKEEPERS**



**LEVERAGING MACHINE
LEARNING FOR POSITIVE
IMPACT**

Exploratory Data Analysis (EDA) of Small Shop
Workers

NAME: Lakshika

ENROLLMENT No : 02901192022

SUBJECT: Machine Learning

SEMESTER:4

OVERVIEW:

UNDERSTANDING THE LANDSCAPE

IN OUR ANALYSIS, WE EMPLOYED THREE DIFFERENT MODELS TO PREDICT AND UNDERSTAND THE BEHAVIOR OF SMALL SHOP WORKERS

THIS DATASET OFFERS A UNIQUE LENS INTO THE WORLD OF SMALL BUSINESS OWNERS, PROVIDING INVALUABLE INSIGHTS INTO THEIR SOCIOECONOMIC LANDSCAPE. WE FOCUS ON PIVOTAL VARIABLES INCLUDING AGE, INCOME, EDUCATION , MORTGAGE, PERSONAL LOAN, AND MORE, OFFERING COMPREHENSIVE VIEW OF THE CHALLENGES AND OPPORTUNITIES FACED BY LOCAL ENTREPRENEURS.

Logistic Regression: This model is suitable for binary classification tasks, making it ideal for predicting outcomes such as whether a small shop worker will take a personal loan or not.

Decision Trees: Decision trees provide a clear and interpretable way to understand the decision-making process based on features such as age, income, and education. They can uncover important relationships between variables and outcomes.

Random Forest: Random forests are an ensemble learning method that combines multiple decision trees to improve predictive accuracy and reduce overfitting. They can handle a large number of input variables and are robust to outliers and noisy data.

For evaluation, we utilized a train-test split methodology, where we trained our models on a subset of the data and evaluated their performance on a separate test set. This helps to ensure that our models generalize well to new, unseen data.

Bank_Personal_Loan_Modelling - Excel (Product Activation Failed)

FileHomeInsertPage LayoutFormulasDataReviewViewTell me what you want to do...

Cut

Copy

Paste

Format Painter

Clipboard

Calibri

11

B

I

U

Font

Wrap Text

Alignment

General

%

Number

Conditional Formatting

Format as Table

Normal

Bad

Neutral

Calculation

Styles

A1

ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	Age	Experience	Income	ZIP Code	Family	CCAvg	Education	Mortgage	Personal L	Securities	CD Accour	Online	CreditCard		
2	1	25	1	49	91107	4	1.6	1	0	0	1	0	0	0		
3	2	45	19	34	90089	3	1.5	1	0	0	1	0	0	0		
4	3	39	15	11	94720	1	1	1	0	0	0	0	0	0		
5	4	35	9	100	94112	1	2.7	2	0	0	0	0	0	0		
6	5	35	8	45	91330	4	1	2	0	0	0	0	0	1		
7	6	37	13	29	92121	4	0.4	2	155	0	0	0	1	0		
8	7	53	27	72	91711	2	1.5	2	0	0	0	0	1	0		
9	8	50	24	22	93943	1	0.3	3	0	0	0	0	0	1		
10	9	35	10	81	90089	3	0.6	2	104	0	0	0	1	0		
11	10	34	9	180	93023	1	8.9	3	0	1	0	0	0	0		
12	11	65	39	105	94710	4	2.4	3	0	0	0	0	0	0		
13	12	29	5	45	90277	3	0.1	2	0	0	0	0	1	0		
14	13	48	23	114	93106	2	3.8	3	0	0	1	0	0	0		
15	14	59	32	40	94920	4	2.5	2	0	0	0	0	1	0		
16	15	67	41	112	91741	1	2	1	0	0	1	0	0	0		
17	16	60	30	22	95054	1	1.5	3	0	0	0	0	1	1		
18	17	38	14	130	95010	4	4.7	3	134	1	0	0	0	0		
19	18	42	18	81	94305	4	2.4	1	0	0	0	0	0	0		
20	19	46	21	193	91604	2	8.1	3	0	1	0	0	0	0		
21	20	55	28	21	94720	1	0.5	2	0	0	1	0	0	1		
22	21	56	31	25	94015	4	0.9	2	111	0	0	0	1	0		
23	22	57	27	63	90095	3	2	3	0	0	0	0	1	0		
24	23	29	5	62	90277	1	1.2	1	260	0	0	0	1	0		
25	24	44	18	43	91320	2	0.7	1	163	0	1	0	0	0		
26	25	36	11	152	95521	2	3.9	1	159	0	0	0	0	1		
27	26	43	19	29	94305	3	0.5	1	97	0	0	0	1	0		
28	27	40	16	83	95064	4	0.2	3	0	0	0	0	0	0		

With over 5000 records and 14 distinct columns, this dataset provides a rich source of information for our analysis.



IMPORTANCE OF PROJECT'S KEY FEATURES

- **Age :** Understanding the age distribution helps identify the demographics of small business owners and their career trajectories.
- **Income :** Income levels provide insights into the financial stability and growth potential of small businesses .
- **Education :** Education levels indicate the skill sets and knowledge base of entrepreneurs , influencing business strategies and decision-making.
- **Mortgage :** Mortgage data reveals the asset ownership and financial obligations of business owners, impacting investment capacity.
- **Personal Loan :** Analysis of personal loan status sheds light on the borrowing behavior and financial risk tolerance of entrepreneurs.
- **Other Variables :** Additional variables such as ZIP code, family size, and credit card usage offer supplementary insights into consumer behavior and market dynamics.



1.Feature Creation:

```
# Checking for missing values
missing_values = data.isnull().sum()
print(missing_values)
# Performing one-hot encoding for categorical
data_encoded = pd.get_dummies(data)

# Display the encoded dataset
print(data_encoded.head())
```

ID	0
Age	0

- Utilized interaction features to capture synergistic effects.
- Generated polynomial features to represent nonlinear relationships.
- Introduced domain-specific features based on industry knowledge.



Exploratory analysis through visualization provided crucial insights into the demographics, socioeconomic status, and potential challenges faced by local workers. By examining variables like age, income, education level, and mortgage status, we gained valuable insights that serve as the foundation for further analysis and model development to support and empower this important segment of the workforce.

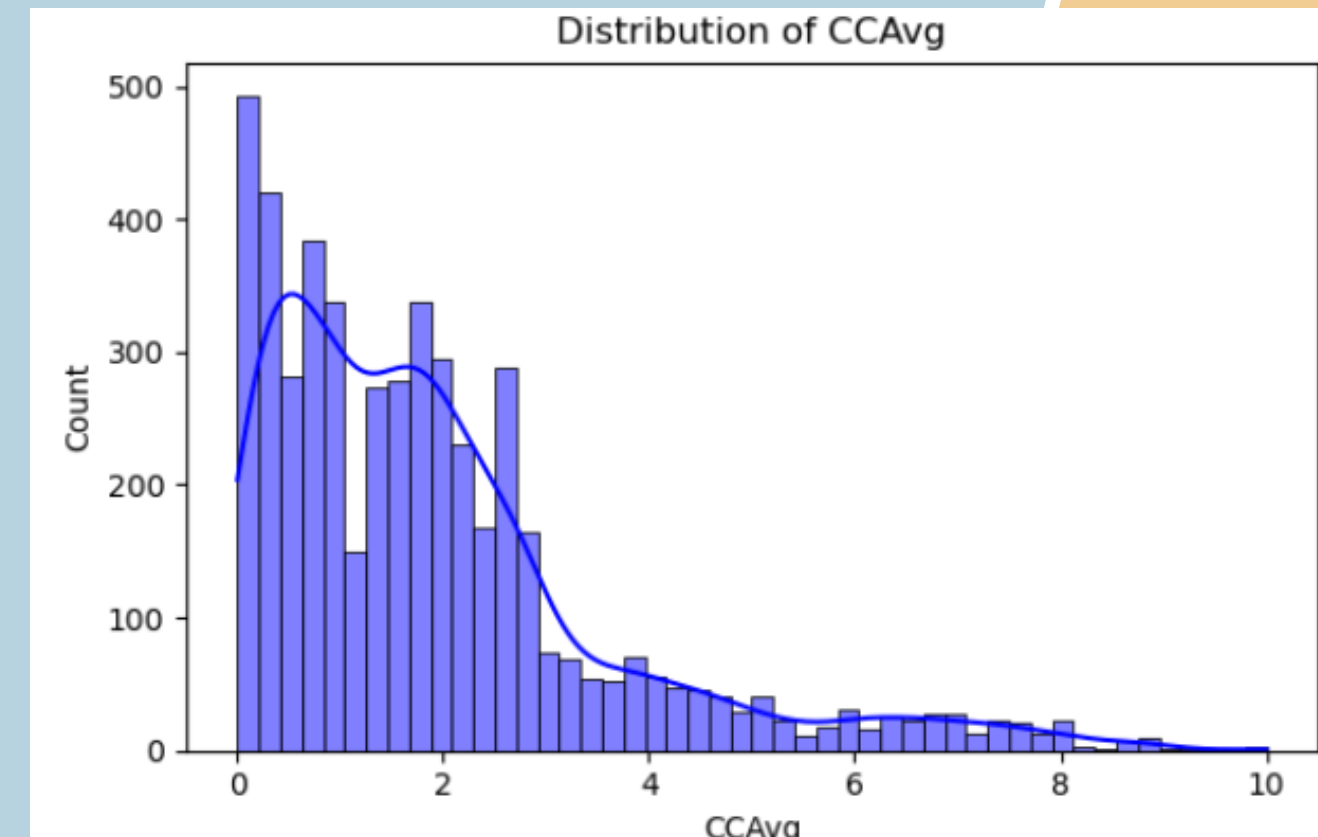
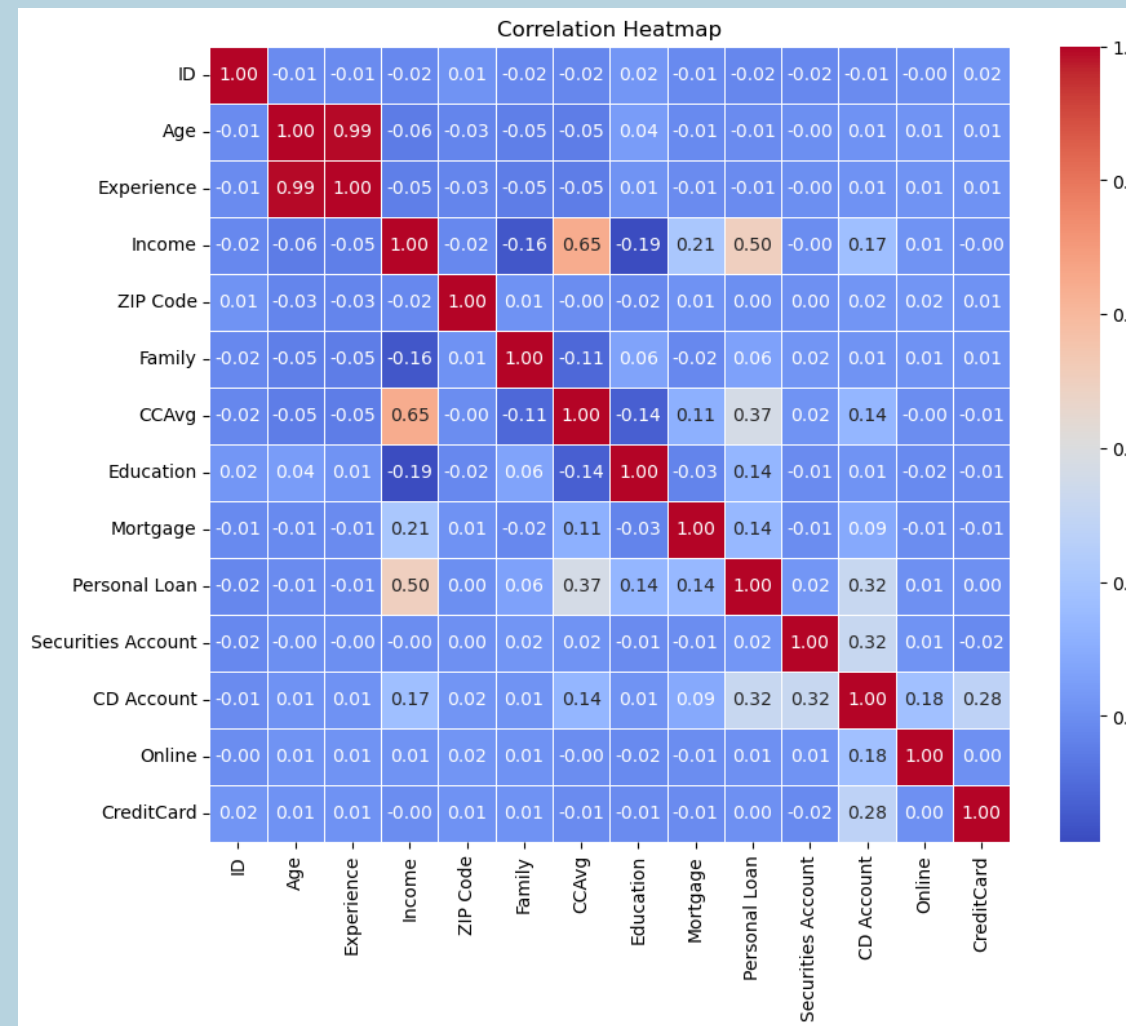
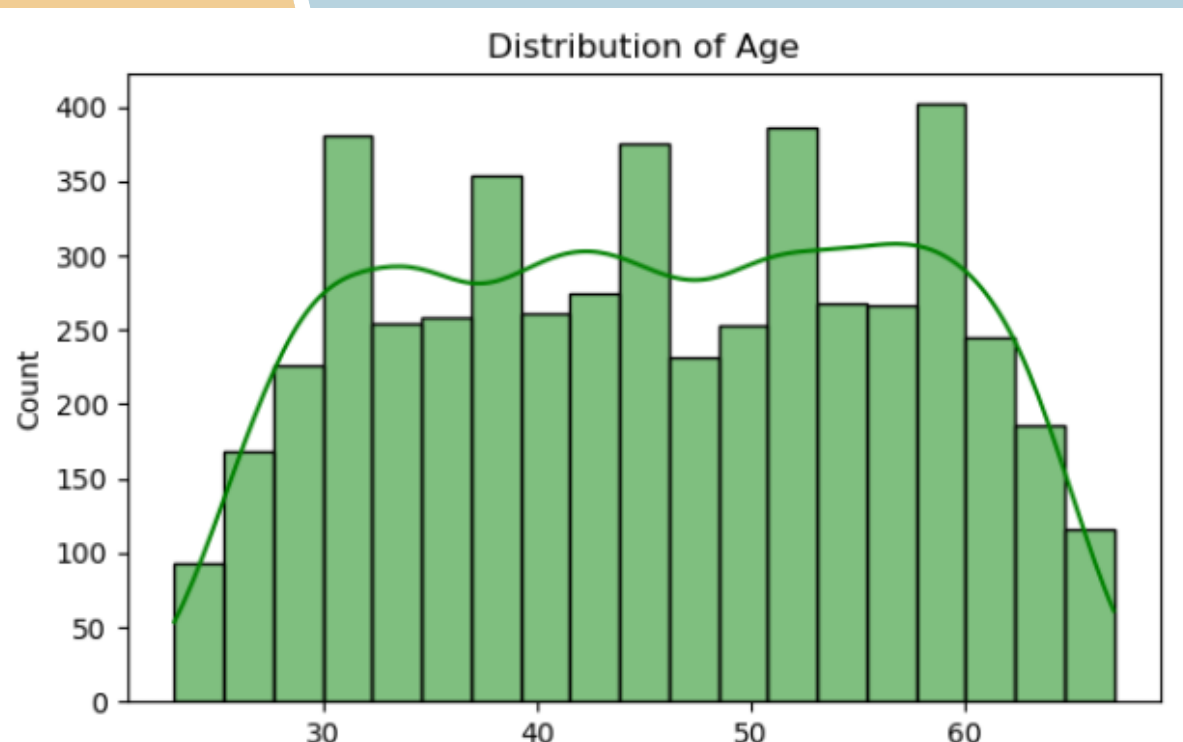
VISUALIZATION AND INSIGHTS GAINED FROM EXPLORATORY ANALYSIS

2.Feature Selection:

- Employed univariate statistical tests for selecting relevant features.
- Utilized Recursive Feature Elimination (RFE) for iterative selection.
- Leveraged model-based feature selection using importance scores.

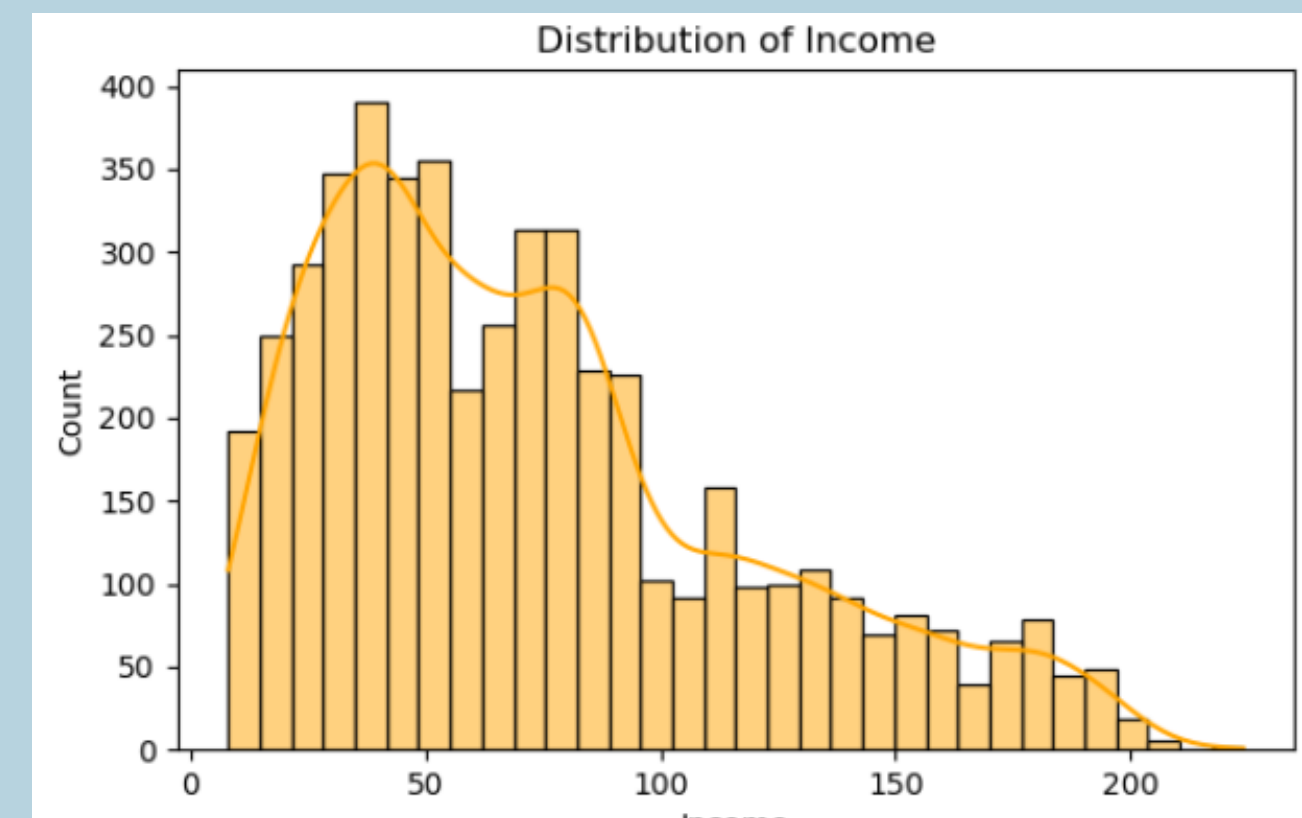
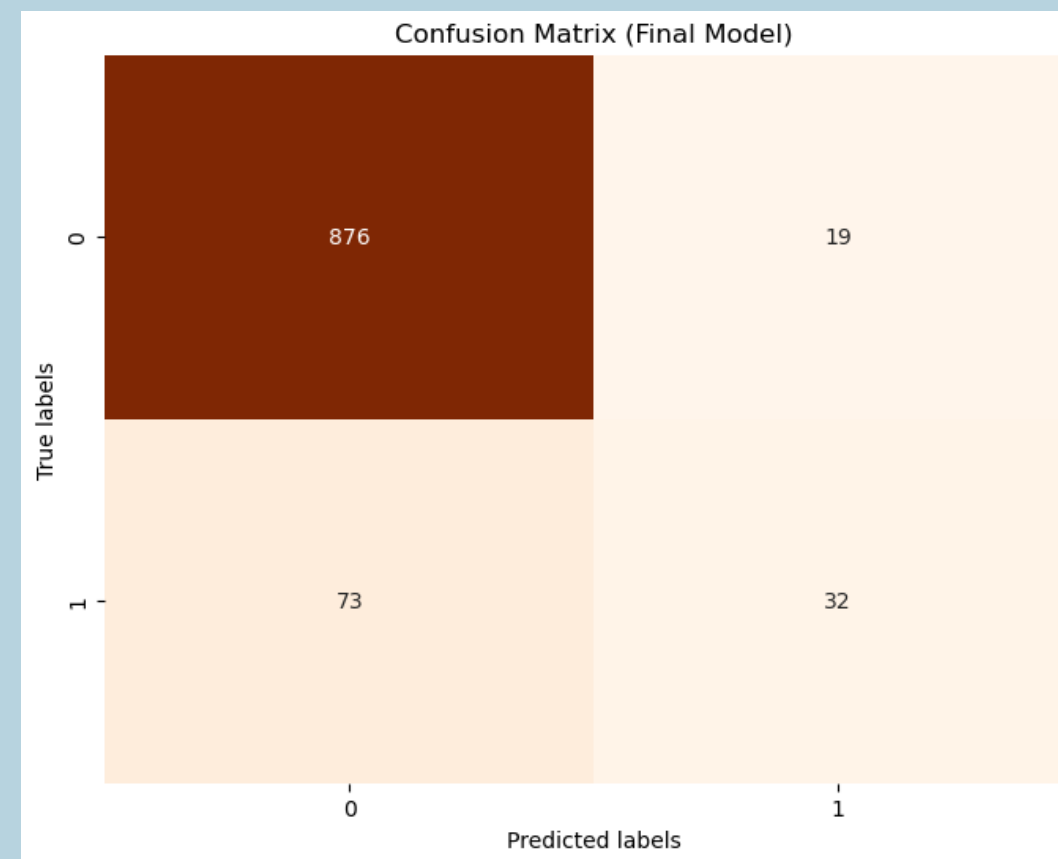
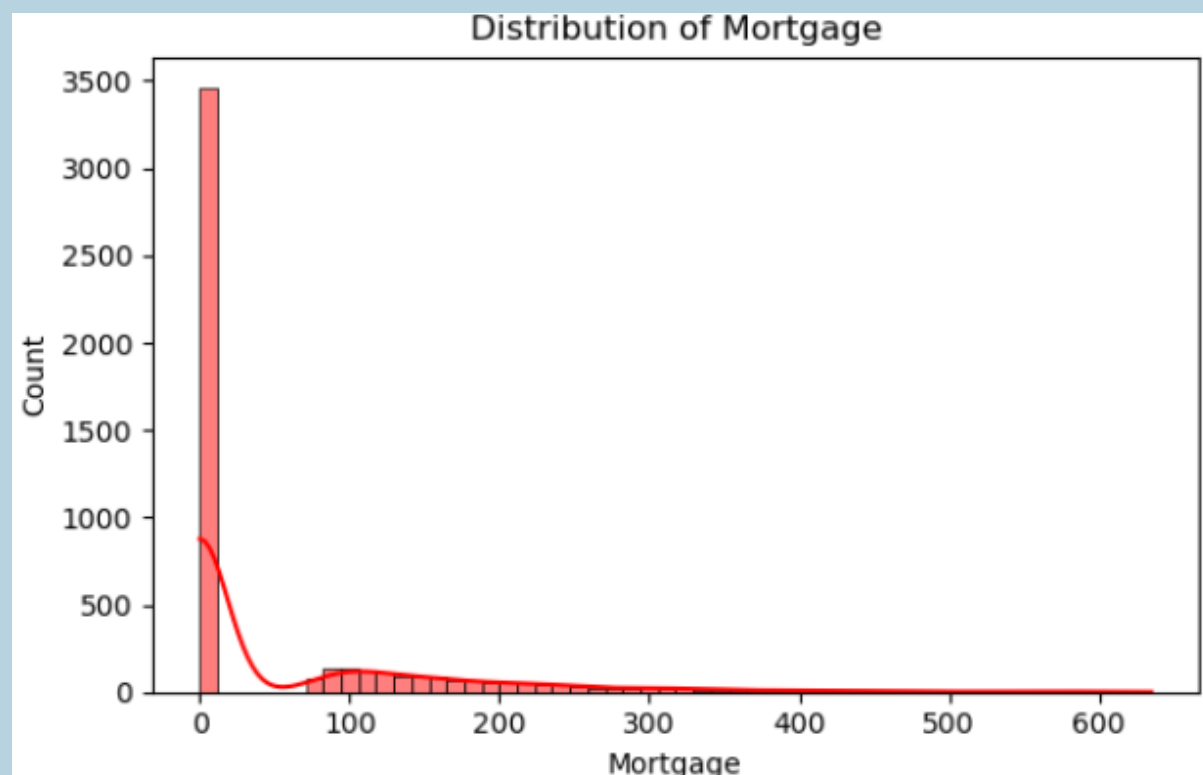


INSIGHTS FROM DATA ANALYSIS



- Histogram:**
- Visualize the distribution of key numerical features like Age, Income, CCAvg, and Mortgage.
 - Identify central tendencies, skewness, and potential outliers.
 - Gain insights into the spread and shape of the data distributions.

- Correlation Heatmap:**
- Provides insights into the pairwise relationships between different variables.



CONCLUSION AND FUTURE SCOPE

In conclusion, our analysis provides valuable insights into the lives of small business owners and highlights opportunities for support and empowerment. Moving forward, we aim to leverage these insights to develop targeted interventions and policies that address the needs of this important segment of our economy. With continued research and collaboration, we can make a meaningful impact on the lives of local entrepreneurs and foster economic growth and prosperity.

1. **Further Research:** Explore additional datasets or conduct more in-depth analysis to gain a deeper understanding of the challenges and opportunities faced by small business owners.
2. **Policy Development:** Work with policymakers and stakeholders to translate your findings into actionable policies and initiatives that support small business growth and sustainability.
3. **Community Engagement:** Engage with local communities and small business owners to gather feedback and input on potential solutions, ensuring that interventions are tailored to their specific needs.
4. **Impact Assessment:** Continuously monitor and evaluate the effectiveness of implemented interventions to measure their impact on local economies and the well-being of small business owners.
5. **Collaboration:** Foster collaboration with other researchers, organizations, and government agencies to leverage expertise and resources for collective efforts aimed at supporting small businesses.

REFERENCES

Dataset:

Code:

THANK YOU

