Property Purchase Analyst

Paksi Yudha Mahendra

ABOUT DATASET

Global Property Purchase

This dataset provides a global property purchase decisions with 200,000 records across 20+ countries and major cities.

Source: link

OBJECTIVE

Determine the estimated loan amount based on the correlation of other data.

ABOUT DATASET (Features)

Property Details:

- Location (countries and major cities)
- Size
- Price
- Type
- Amenities

Financial Information:

- Salary
- Loan
- EMI (Equated Monthly Installments)
- Expenses breakdown

ABOUT DATASET (Features)

Subjective Ratings

- Satisfaction scores
- Neighborhood ratings
- Connectivity assessments

Target Variable

- Decision → Buy (1) / Not Buy (0)

METHODOLOGY

- Data Cleaning
- Uji Correlation (Spearman and Pearson)
- Chi Square
- Regresi Linear

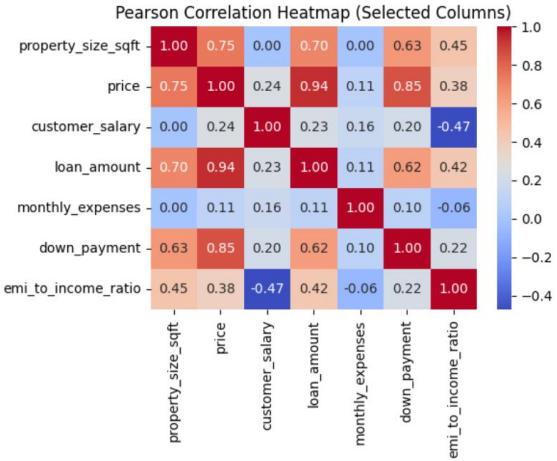
Data Cleaning

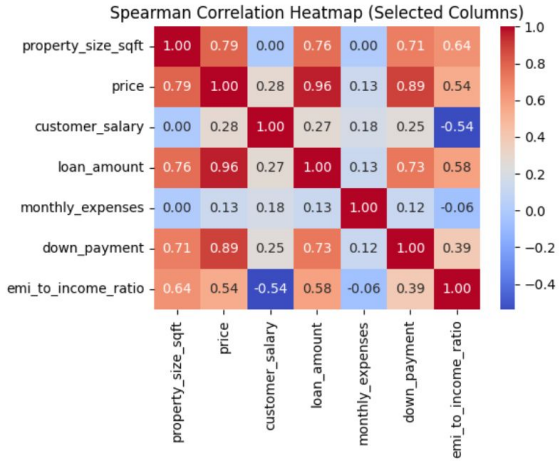
```
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   property_id               200000 non-null  int64
 1   country                   200000 non-null  object
 2   city                      200000 non-null  object
 3   property_type             200000 non-null  object
 4   furnishing_status         200000 non-null  object
 5   property_size_sqft        200000 non-null  int64
 6   price                     200000 non-null  int64
 7   constructed_year          200000 non-null  int64
 8   previous_owners           200000 non-null  int64
 9   rooms                     200000 non-null  int64
 10  bathrooms                 200000 non-null  int64
 11  garage                    200000 non-null  int64
 12  garden                    200000 non-null  int64
 13  crime_cases_reported      200000 non-null  int64
 14  legal_cases_on_property   200000 non-null  int64
 15  customer_salary           200000 non-null  int64
 16  loan_amount               200000 non-null  int64
 17  loan_tenure_years         200000 non-null  int64
 18  monthly_expenses          200000 non-null  int64
 19  down_payment              200000 non-null  int64
 20  emi_to_income_ratio       200000 non-null  float64
 21  satisfaction_score        200000 non-null  int64
 22  neighbourhood_rating      200000 non-null  int64
 23  connectivity_score        200000 non-null  int64
 24  decision                  200000 non-null  int64
dtypes: float64(1), int64(20), object(4)
memory usage: 38.1+ MB
```

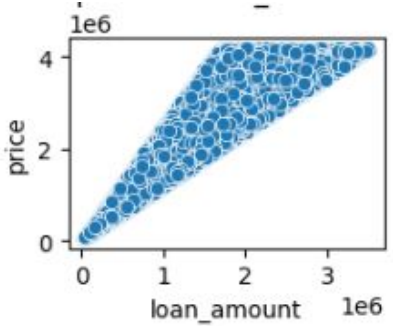
At this stage, a check is performed to see if there is any incomplete data and data formatting, such as changing the year data from a string to dt.year.

```
1 # convert year data to
2 rawdata['constructed_year'] = pd.to_datetime(rawdata['constructed_year'], format='%Y', errors='coerce').dt.year
3
```

Uji Correlation (Spearman and Pearson)







Loan Amount has a strong correlation with property price.

We will use these two variables to create a model for predicting loan amount based on property price.

Chi Square

Pivot Table: Count of Property Types by Furnishing Status

furnishing_status	Fully-Furnished	Semi-Furnished	Unfurnished
property_type			
Apartment	11109	11195	11094
Farmhouse	11234	11184	11100
Independent House	11169	10927	11238
Studio	10924	11164	10920
Townhouse	11219	11190	10986
Villa	11174	11013	11160

Since this dataset contains various types of properties and furnishing statuses, we will test it with Chi Square to see if there is a significant relationship between the two variables to determine the linear regression model.

Chi Square

Contingency Table:

furnishing_status	Fully-Furnished	Semi-Furnished	Unfurnished
property_type			
Apartment	11109	11195	11094
Farmhouse	11234	11184	11100
Independent House	11169	10927	11238
Studio	10924	11164	10920
Townhouse	11219	11190	10986
Villa	11174	11013	11160

p value is 0.2118255659841176
Independent (H0 holds true)

From these results, it can be concluded that there is no significant relationship between property type and furnishing status.

So we can directly create a model using linear regression.

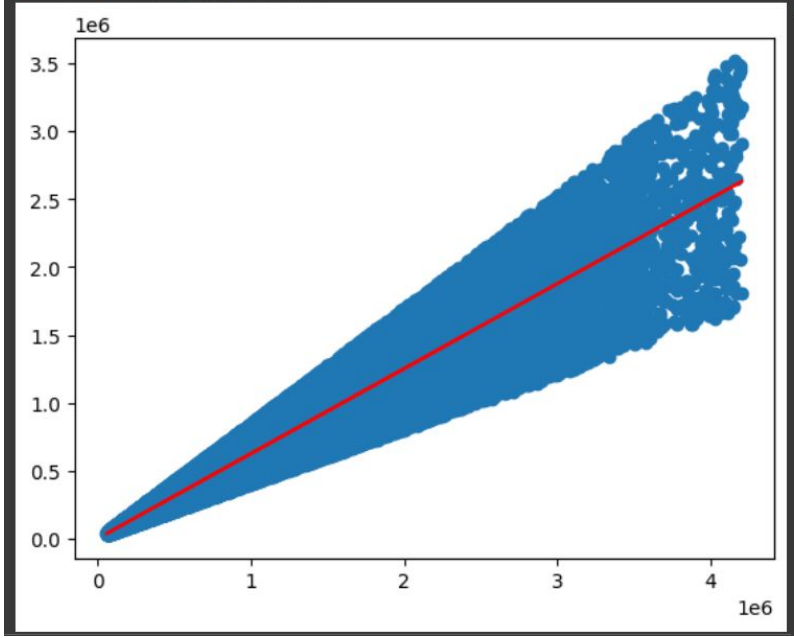
Regresi Linear

We will create models for each type of property in the dataset, including:

- Farmhouse
- Apartment
- Townhouse
- Villa
- Studio
- Independent

Regresi Linear (Farmhouse & Apartment)

```
Coefficient: 0.6269581059704709
Intercept: -2081.0332947318675
R-squared: 0.8809488947998455
```




```
Apartment Property Type:
Coefficient: 0.6242791871039671
Intercept: 453.0869308139663
R-squared: 0.8797730572623043
```

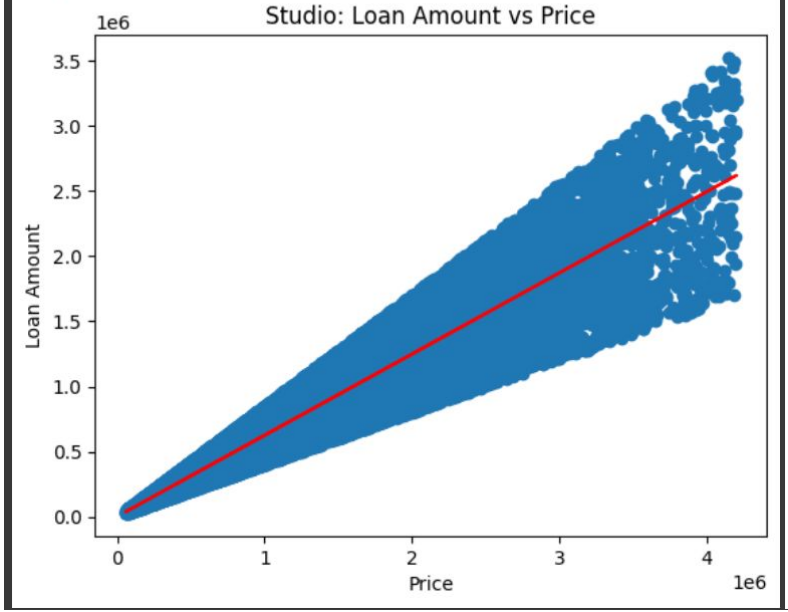
Regresi Linear (Townhouse & Villa)

```
Townhouse Property Type:
Coefficient: 0.6241543317473038
Intercept: 395.9454249413684
R-squared: 0.8795903032996141
```


```
Villa Property Type:
Coefficient: 0.625584367643792
Intercept: -456.09043003746774
R-squared: 0.8807482916313163
```

Regresi Linear (Studio & Independent House)

```
Studio Property Type:
Coefficient: 0.6221433569071175
Intercept: 3054.430194844841
R-squared: 0.8769493642752625
```




```
Independent House Property Type:
Coefficient: 0.6266664865166969
Intercept: -188.86981956462841
R-squared: 0.8782331103356553
```

Regresi Linear

The average R-square value for the linear regression model created is 80%, which means that the model can be used to predict the loan amount based on the price for each type of property.

Regresi Linear

For example,

if an independent house is priced at 4,750,000, what is the loan amount?

Loan_amount = -188,87 + 0.62 * (price)

Loan_amount = -188,87 + 0.62 * (4.750.000)

Loan_amount = 2.944.811,13

THANK YOU

Link Google Colab