01-credit-risk-model-feature-engineering

September 22, 2025

1 Model Development with FEAST Feature Store Tutorial

1.1 Objective

This three-part series of Notebooks is to demostrate how FEAST is integrated as the ML Feature Store during the Model Development Journey:

1. 01 - Feature Engineering
2. 02 - Feature Registration
3. 03 - Feature Consumption (Model Training)

1.2 Model

We develop a model that predicts if a credit loan customer can be a bad credit loan using the credit risk dataset.

1.3 Dataset

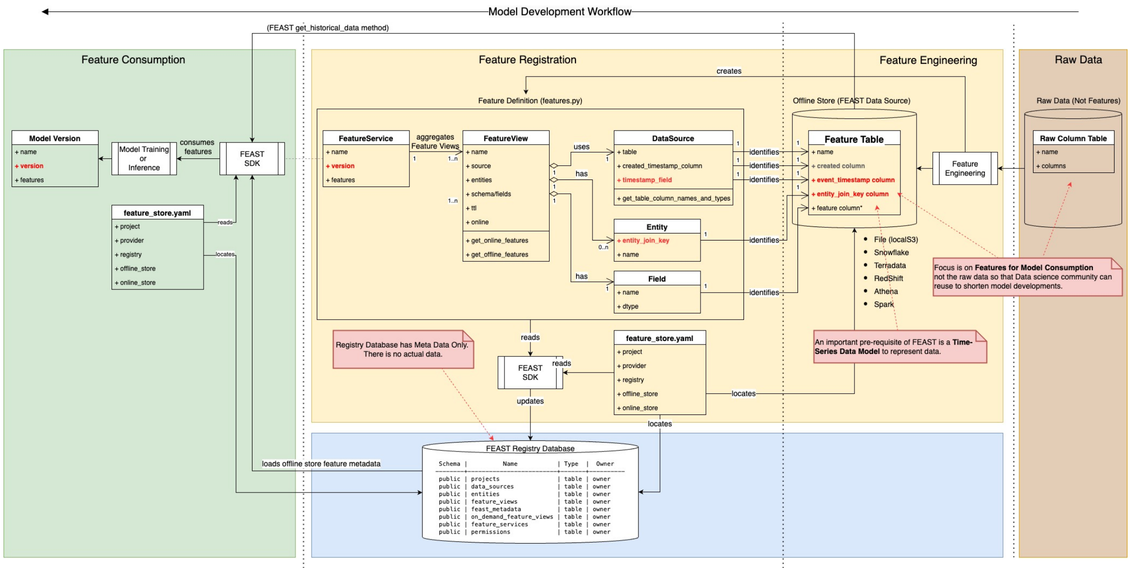
Use the trimmed Bank Customer Credit Risk dataset to have only:

Age (numeric) Sex (text: male, female) Job (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled) Housing (text: own, rent, or free) Saving accounts (text - little, moderate, quite rich, rich) Checking account (numeric, in DM - Deutsch Mark) Credit amount (numeric, in DM) Duration (numeric, in month) Purpose(text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others Risk (Value target - Good or Bad Risk)

See german-credit-data-with-risk to see the data in Kaggle.

1.4 Model Development Workflow (Right to Left)

The diagram depicts the model development flow and how FEAST is involved to provide the Feature Store functionality.



---

2 Local Deployment Environment

The tutorial provides the automated script to setup the environment to run the model development workflow.

---

3 Feature Engineering

The goal is **Features in the Offline Store** for Model Training Consumption.

3.1 This Notebook (01 Feature Engineering)

This notebook goes through the steps of:

1. Data Exploration,
2. Feature Engineering to generate the **Features to be consumed by the model training**, and,
3. Upload them to the **Offline Store**.

---

4 Setup

4.1 Note

Install OpenMP runtime on Macbook.

```
brew install libomp
```

5 Raw Data (Bank Customer Credit Risk)

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               1000 non-null   int64
 1   Sex               1000 non-null   object
 2   Job               1000 non-null   int64
 3   Housing           1000 non-null   object
 4   Saving accounts   817 non-null    object
 5   Checking account  606 non-null    object
 6   Credit amount     1000 non-null   int64
 7   Duration          1000 non-null   int64
 8   Purpose           1000 non-null   object
 9   Risk              1000 non-null   float64
dtypes: float64(1), int64(4), object(5)
memory usage: 85.9+ KB
None
```

```
   Age     Sex  Job Housing Saving accounts Checking account  Credit amount
 Duration              Purpose  Risk
 0  67    male    2     own             NaN           little           1169
 6            radio/TV   0.0
 1  22  female    2     own          little         moderate           5951
 48            radio/TV   1.0
 2  49    male    1     own          little              NaN           2096
 12            education   0.0
 3  45    male    2    free          little           little           7882
 42  furniture/equipment   0.0
 4  53    male    2    free          little           little           4870
 24                  car   1.0
```

5.0.1 Column Types

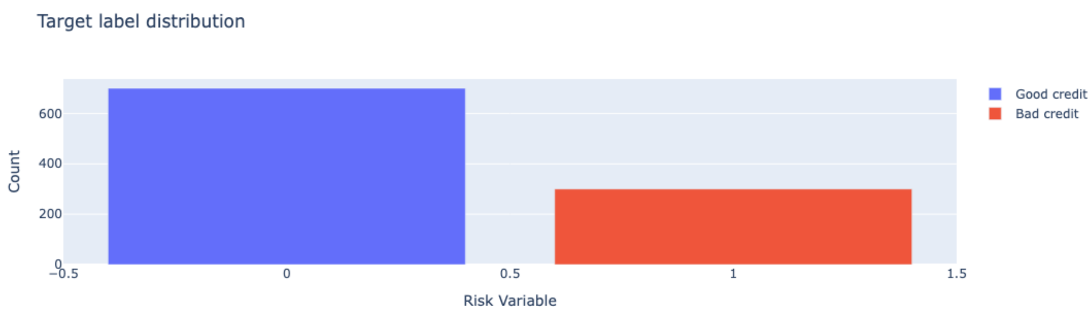
5.0.2 Unique Categorical Values

```
Sex              : [male, female]
Job              : [2, 1, 3, 0]
Housing          : [own, free, rent]
Saving accounts  : [little, quite rich, rich, moderate]
Checking account : [little, moderate, rich]
Purpose          : [radio/TV, education, furniture/equipment, car, business,
domestic appliances, repairs, vacation/others]
```

6 EDA

6.1 Target label (Risk) and imbalance

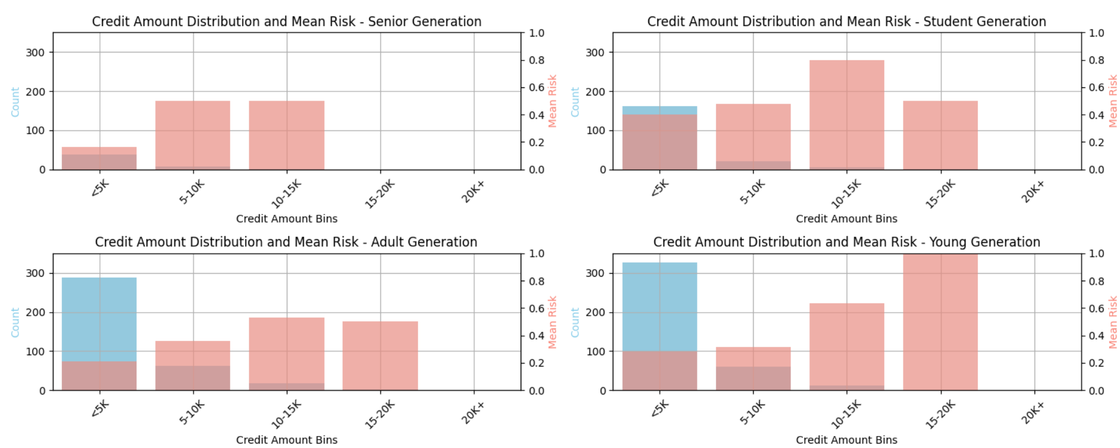
```
   Age   Sex  Job Housing Saving accounts Checking account  Credit amount
 Duration   Purpose  Risk Generation Amount
 0   67  male    2     own             NaN           little           1169
 6  radio/TV   0.0     Senior    <5K
```




```
Non Risky data is [2.33] more than Risky.
```

6.2 Risk by Credit Amount

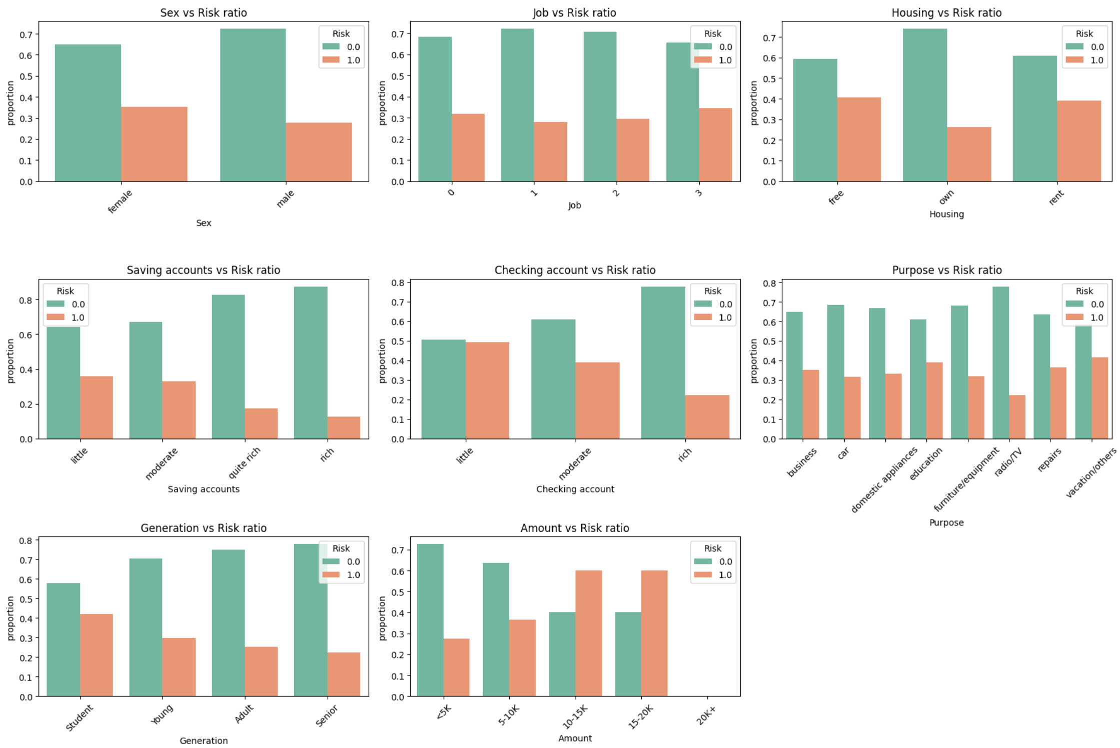
Higher the amount, higher the risk.

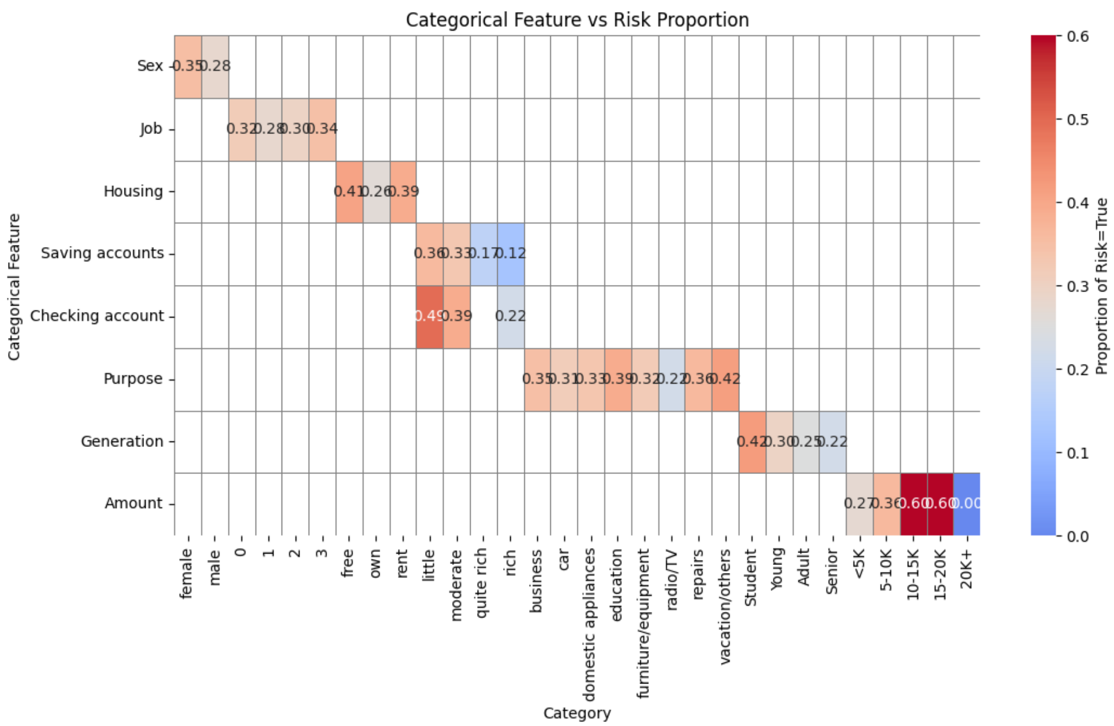


7 Risk Correlation

```
INFO:matplotlib.category:Using categorical units to plot a list of strings that
are all parsable as floats or dates. If these strings should be plotted as
```

```
numbers, cast to the appropriate data type before plotting.
INFO:matplotlib.category:Using categorical units to plot a list of strings that
are all parsable as floats or dates. If these strings should be plotted as
numbers, cast to the appropriate data type before plotting.
```





---

8 Featur Engineering

Supposing the EDA has been done and concluded:

1. Not to use `Duration` column.
2. Convert all other columns to categorical.
3. One Hot Encode all the categorical columns.

9 Features for Model Consumption

Verify the features to be consumed by the Model Training

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, 0 to 999
Data columns (total 36 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   risk                            1000 non-null   float64
 1   gender_female                   1000 non-null   float32
 2   gender_male                     1000 non-null   float32
 3   job_0                           1000 non-null   float32
 4   job_1                           1000 non-null   float32
```

```
 5   job_2                        1000 non-null   float32
 6   job_3                        1000 non-null   float32
 7   housing_free                 1000 non-null   float32
 8   housing_own                  1000 non-null   float32
 9   housing_rent                 1000 non-null   float32
 10  saving_accounts_little       1000 non-null   float32
 11  saving_accounts_moderate     1000 non-null   float32
 12  saving_accounts_no_inf       1000 non-null   float32
 13  saving_accounts_quite_rich   1000 non-null   float32
 14  saving_accounts_rich         1000 non-null   float32
 15  checking_account_little      1000 non-null   float32
 16  checking_account_moderate    1000 non-null   float32
 17  checking_account_no_inf      1000 non-null   float32
 18  checking_account_rich        1000 non-null   float32
 19  purpose_business             1000 non-null   float32
 20  purpose_car                  1000 non-null   float32
 21  purpose_domestic_appliances  1000 non-null   float32
 22  purpose_education            1000 non-null   float32
 23  purpose_furniture_equipment  1000 non-null   float32
 24  purpose_radio_tv             1000 non-null   float32
 25  purpose_repairs              1000 non-null   float32
 26  purpose_vacation_others      1000 non-null   float32
 27  generation_student           1000 non-null   float32
 28  generation_young             1000 non-null   float32
 29  generation_adult             1000 non-null   float32
 30  generation_senior            1000 non-null   float32
 31  amount_0                     1000 non-null   float32
 32  amount_1                     1000 non-null   float32
 33  amount_2                     1000 non-null   float32
 34  amount_3                     1000 non-null   float32
 35  amount_4                     1000 non-null   float32
dtypes: float32(35), float64(1)
memory usage: 152.3 KB

   risk  gender_female  gender_male  job_0  job_1  job_2  job_3  housing_free
 housing_own  housing_rent  saving_accounts_little  saving_accounts_moderate
 saving_accounts_no_inf  saving_accounts_quite_rich  saving_accounts_rich
 checking_account_little  checking_account_moderate  checking_account_no_inf
 checking_account_rich  purpose_business  purpose_car
 purpose_domestic_appliances  purpose_education  purpose_furniture_equipment
 purpose_radio_tv  purpose_repairs  purpose_vacation_others  generation_student
 generation_young  generation_adult  generation_senior  amount_0  amount_1
 amount_2  amount_3  amount_4
 0   0.0            0.0          1.0    0.0    0.0    1.0    0.0           0.0
 1.0           0.0                     0.0                       0.0
 1.0                         0.0                     0.0                    1.0
 0.0                      0.0                      0.0               0.0
```

```
0.0                          0.0                0.0                            0.0
1.0              0.0                        0.0                  0.0
0.0               0.0                1.0       1.0       0.0       0.0       0.0
0.0
1   1.0            1.0          0.0    0.0    0.0    1.0    0.0           0.0
1.0           0.0                     1.0                        0.0
0.0                         0.0                    0.0                       0.0
1.0                     0.0                    0.0                0.0
0.0                          0.0                0.0                            0.0
1.0              0.0                        0.0                  1.0
0.0               0.0                0.0       0.0       1.0       0.0       0.0
0.0
2   0.0            0.0          1.0    0.0    1.0    0.0    0.0           0.0
1.0           0.0                     1.0                        0.0
0.0                         0.0                    0.0                       0.0
0.0                     1.0                    0.0                0.0
0.0                          0.0                1.0                            0.0
0.0              0.0                        0.0                  0.0
0.0               1.0                0.0       1.0       0.0       0.0       0.0
0.0
```

9.1 Save Features to csv

9.2 Dummpy fields for FEAST

FEAST requires time-series data structure and an entity key of each time-series record.

```
   risk  gender_female  gender_male  job_0  job_1  job_2  job_3  housing_free
housing_own  housing_rent  saving_accounts_little  saving_accounts_moderate
saving_accounts_no_inf  saving_accounts_quite_rich  saving_accounts_rich
checking_account_little  checking_account_moderate  checking_account_no_inf
checking_account_rich  purpose_business  purpose_car
purpose_domestic_appliances  purpose_education  purpose_furniture_equipment
purpose_radio_tv  purpose_repairs  purpose_vacation_others  generation_student
generation_young  generation_adult  generation_senior  amount_0  amount_1
amount_2  amount_3  amount_4            event_timestamp
created  entity_id
0   0.0            0.0          1.0    0.0    0.0    1.0    0.0           0.0
1.0           0.0                     0.0                        0.0
1.0                         0.0                    0.0                       1.0
0.0                     0.0                    0.0                0.0
0.0                          0.0                0.0                            0.0
1.0              0.0                        0.0                  0.0
0.0               0.0                1.0       1.0       0.0       0.0       0.0
0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          1
1   1.0            1.0          0.0    0.0    0.0    1.0    0.0           0.0
1.0           0.0                     1.0                        0.0
0.0                         0.0                    0.0                       0.0
```

```
1.0                      0.0                     0.0                 0.0
0.0                          0.0                 0.0                           0.0
1.0              0.0                       0.0                   1.0
0.0               0.0                0.0       0.0       1.0       0.0       0.0
0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          2
2   0.0            0.0          1.0    0.0    1.0    0.0    0.0           0.0
1.0           0.0                    1.0                     0.0
0.0                         0.0                    0.0                    0.0
0.0                      1.0                     0.0                 0.0
0.0                          0.0                 1.0                           0.0
0.0              0.0                       0.0                   0.0
0.0               1.0                0.0       1.0       0.0       0.0       0.0
0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          3
3   0.0            0.0          1.0    0.0    0.0    1.0    0.0           1.0
0.0           0.0                    1.0                     0.0
0.0                         0.0                    0.0                    1.0
0.0                      0.0                     0.0                 0.0
0.0                          0.0                 0.0                           1.0
0.0              0.0                       0.0                   0.0
0.0               1.0                0.0       0.0       1.0       0.0       0.0
0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          4
4   1.0            0.0          1.0    0.0    0.0    1.0    0.0           1.0
0.0           0.0                    1.0                     0.0
0.0                         0.0                    0.0                    1.0
0.0                      0.0                     0.0                 0.0
1.0                          0.0                 0.0                           0.0
0.0              0.0                       0.0                   0.0
0.0               1.0                0.0       1.0       0.0       0.0       0.0
0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          5
```

---

10 Features to Offline Store

10.1 Offline Store Paramters

```
INFO:root:Successfully loaded YAML:
../deployment/feast/feature_repository/feature_store.yaml
```

```
{'host': 'localhost',
 'port': 5432,
 'database': 'offline_features',
 'user': 'dbadm',
 'schema': 'credit',
 'table_name': 'customer_credit_risk_offline_features'}
```

10.2 Insert into offline store table

```
INFO:root:Table 'credit.customer_credit_risk_offline_features' does not exist
INFO:root:Available tables: []
Engine auto-disposed
INFO:root:Inserting [1000] rows in [1] batches…
```

```
Inserting batches: 100%|

        | 1/1 [00:00<00:00,  2.09it/s]
```

```
Successfully inserted all 1000 rows!
Engine auto-disposed
```

10.3 Verify

```
Engine auto-disposed
```

```
risk
gender_female
gender_male
job_0
job_1
job_2
job_3
housing_free
housing_own
housing_rent
saving_accounts_little
saving_accounts_moderate
saving_accounts_no_inf
saving_accounts_quite_rich
saving_accounts_rich
checking_account_little
checking_account_moderate
checking_account_no_inf
checking_account_rich
purpose_business
purpose_car
purpose_domestic_appliances
purpose_education
purpose_furniture_equipment
purpose_radio_tv
purpose_repairs
purpose_vacation_others
generation_student
generation_young
generation_adult
```

```
generation_senior
amount_0
amount_1
amount_2
amount_3
amount_4
event_timestamp
created
entity_id

   risk  gender_female  gender_male  job_0  job_1  job_2  job_3  housing_free
 housing_own  housing_rent  saving_accounts_little  saving_accounts_moderate
 saving_accounts_no_inf  saving_accounts_quite_rich  saving_accounts_rich
 checking_account_little  checking_account_moderate  checking_account_no_inf
 checking_account_rich  purpose_business  purpose_car
 purpose_domestic_appliances  purpose_education  purpose_furniture_equipment
 purpose_radio_tv  purpose_repairs  purpose_vacation_others  generation_student
 generation_young  generation_adult  generation_senior  amount_0  amount_1
 amount_2  amount_3  amount_4            event_timestamp
 created  entity_id
 0   0.0            0.0           1.0    0.0    0.0    1.0    0.0           0.0
 1.0           0.0                     0.0                       0.0
 1.0                         0.0                  0.0                       1.0
 0.0                      0.0                    0.0               0.0
 0.0                          0.0                0.0                          0.0
 1.0              0.0                     0.0                  0.0
 0.0               0.0                 1.0       1.0       0.0       0.0       0.0
 0.0 2025-09-18 12:27:23.123026 2025-09-18 12:27:23.123026          1
```