1lmfiefys

March 31, 2025

1 Experiment Notebook

1.1 0. Setup Environment

1.1.1 0.a Install Mandatory Packages

Do not modify this code before running it

[1]:
```
# Do not modify this code

import os
import sys
from pathlib import Path

COURSE = "36106"
ASSIGNMENT = "AT1"
DATA = "data"

asgmt_path = f"{COURSE}/assignment/{ASSIGNMENT}"
root_path = "./"

print("###### Install required Python packages ######")
! pip install -r https://raw.githubusercontent.com/aso-uts/labs_datasets/main/
  ↪36106-mlaa/requirements.txt

if os.getenv("COLAB_RELEASE_TAG"):

    from google.colab import drive
    from pathlib import Path

    print("\n###### Connect to personal Google Drive ######")
    gdrive_path = "/content/gdrive"
    drive.mount(gdrive_path)
    root_path = f"{gdrive_path}/MyDrive/"

print("\n###### Setting up folders ######")
folder_path = Path(f"{root_path}/{asgmt_path}/") / DATA
```

```
folder_path.mkdir(parents=True, exist_ok=True)
print(f"\nYou can now save your data files in: {folder_path}")

if os.getenv("COLAB_RELEASE_TAG"):
    %cd {folder_path}
```

```
###### Install required Python packages ######
Requirement already satisfied: pandas==2.2.2 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from -r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2.2.2)
Requirement already satisfied: scikit-learn==1.6.1 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from -r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (1.6.1)
Requirement already satisfied: altair==5.5.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from -r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (5.5.0)
Requirement already satisfied: numpy>=1.23.2 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from pandas==2.2.2->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (1.24.3)
Requirement already satisfied: python-dateutil>=2.8.2 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from pandas==2.2.2->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from pandas==2.2.2->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from pandas==2.2.2->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2023.3)
Requirement already satisfied: scipy>=1.6.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from scikit-
learn==1.6.1->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (1.11.1)
Requirement already satisfied: joblib>=1.2.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from scikit-
learn==1.6.1->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (1.2.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from scikit-
learn==1.6.1->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (3.5.0)
```

```
Requirement already satisfied: jinja2 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (3.1.2)
Requirement already satisfied: jsonschema>=3.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (4.17.3)
Requirement already satisfied: narwhals>=1.14.2 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (1.31.0)
Requirement already satisfied: packaging in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (23.1)
Requirement already satisfied: typing-extensions>=4.10.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (4.12.2)
Requirement already satisfied: attrs>=17.4.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from
jsonschema>=3.0->altair==5.5.0->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (22.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from
jsonschema>=3.0->altair==5.5.0->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (0.18.0)
Requirement already satisfied: six>=1.5 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from python-
dateutil>=2.8.2->pandas==2.2.2->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (1.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/Users/ratikpant/anaconda3/lib/python3.11/site-packages (from
jinja2->altair==5.5.0->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (2.1.1)

###### Setting up folders ######

You can now save your data files in: 36106/assignment/AT1/data
```

1.1.2 0.b Disable Warnings Messages

Do not modify this code before running it

[2]:
```
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

1.1.3 0.c Install Additional Packages

If you are using additional packages, you need to install them here using the command:
`! pip install <package_name>`

[3]:
```
# <Student to fill this section>
```

1.1.4 0.d Import Packages

[4]:
```
import ipywidgets as widgets
import pandas as pd
import altair as alt
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

```
/Users/ratikpant/anaconda3/lib/python3.11/site-
packages/pandas/core/arrays/masked.py:60: UserWarning: Pandas requires version
'1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).
  from pandas.core import (
```

1.2 A. Project Description

[5]:
```
# @title Student Information
wgt_student_name = widgets.Text(
    value=None,
    placeholder='<student to fill this section>',
    description='Student Name:',
    style={'description_width': 'initial'},
    disabled=False
)

wgt_student_id = widgets.Text(
    value=None,
    placeholder='<student to fill this section>',
    description='Student Id:',
    style={'description_width': 'initial'},
    disabled=False
)

widgets.HBox([wgt_student_name, wgt_student_id])
```

[5]:
```
HBox(children=(Text(value='', description='Student Name:', placeholder='<student
to fill this section>', style…
```

[6]:
```
# @title Experiment ID

wgt_experiment_id = widgets.BoundedIntText(
    value=2,
    min=0,
    max=3,
    step=1,
    description='Experiment ID:',
    style={'description_width': 'initial'},
    disabled=False
)
wgt_experiment_id
```

[6]:
```
BoundedIntText(value=2, description='Experiment ID:', max=3,
style=DescriptionStyle(description_width='initial…
```

[7]:
```
# @title Business Objective

wgt_business_objective = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Business Objective:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_business_objective
```

[7]:
```
Textarea(value='', description='Business Objective:',
layout=Layout(height='100%', width='auto'), placeholder=…
```

1.3 B. Experiment Description

[8]:
```
# @title Experiment Hypothesis

wgt_experiment_hypothesis = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Experiment Hypothesis:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_experiment_hypothesis
```

[8]:
```
Textarea(value='', description='Experiment Hypothesis:',
layout=Layout(height='100%', width='auto'), placehold…
```

[9]:
```
# @title Experiment Expectations

wgt_experiment_expectations = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Experiment Expectations:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_experiment_expectations
```

[9]:
```
Textarea(value='', description='Experiment Expectations:',
layout=Layout(height='100%', width='auto'), placeho…
```

1.4 C. Data Understanding

1.4.1 C.1 Load Datasets

Do not change this code

[10]:
```
# Load training data

X_train = pd.read_csv('/Users/ratikpant/Desktop/machine learning/ X_train.csv')
y_train = pd.read_csv('/Users/ratikpant/Desktop/machine learning/ y_train.csv')
```

[11]:
```
# Load validation data
X_val = pd.read_csv('/Users/ratikpant/Desktop/machine learning/ X_val.csv')
y_val = pd.read_csv('/Users/ratikpant/Desktop/machine learning/ y_val.csv')
```

[12]:
```
# Load testing data
X_test = pd.read_csv('/Users/ratikpant/Desktop/machine learning/X_test.csv')
y_test = pd.read_csv('/Users/ratikpant/Desktop/machine learning/y_test.csv')
```

[13]:
```
X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3316 entries, 0 to 3315
Data columns (total 19 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   number_of_bedrooms                    3316 non-null   int64
 1   floor_area                            3316 non-null   int64
 2   current_level                         3316 non-null   float64
```

```
 3   total_level                          3316 non-null   float64
 4   number_of_bathrooms                  3316 non-null   int64
 5   advertised_month                     3316 non-null   int64
 6   average_rent_bath&bed                3316 non-null   float64
 7   suburb_Adelaide                      3316 non-null   int64
 8   suburb_Brisbane                      3316 non-null   int64
 9   suburb_Canberra                      3316 non-null   int64
 10  suburb_Melbourne                     3316 non-null   int64
 11  suburb_Perth                         3316 non-null   int64
 12  suburb_Sydney                        3316 non-null   int64
 13  furnished_Furnished                  3316 non-null   int64
 14  furnished_Semi-Furnished             3316 non-null   int64
 15  furnished_Unfurnished                3316 non-null   int64
 16  tenancy_preference_Bachelors         3316 non-null   int64
 17  tenancy_preference_Bachelors/Family  3316 non-null   int64
 18  tenancy_preference_Family            3316 non-null   int64
dtypes: float64(3), int64(16)
memory usage: 492.3 KB
```

[14]:
```
#changing advertised_month into one hot encoding
```

[15]:
```
X_train = pd.get_dummies(X_train, columns = ['advertised_month'], dtype =int)
```

[16]:
```
X_val.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 983 entries, 0 to 982
Data columns (total 19 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   number_of_bedrooms                   983 non-null    int64
 1   floor_area                           983 non-null    int64
 2   current_level                        983 non-null    float64
 3   total_level                          983 non-null    float64
 4   number_of_bathrooms                  983 non-null    int64
 5   advertised_month                     983 non-null    int64
 6   average_rent_bath&bed                983 non-null    float64
 7   suburb_Adelaide                      983 non-null    int64
 8   suburb_Brisbane                      983 non-null    int64
 9   suburb_Canberra                      983 non-null    int64
 10  suburb_Melbourne                     983 non-null    int64
 11  suburb_Perth                         983 non-null    int64
 12  suburb_Sydney                        983 non-null    int64
 13  furnished_Furnished                  983 non-null    int64
 14  furnished_Semi-Furnished             983 non-null    int64
 15  furnished_Unfurnished                983 non-null    int64
 16  tenancy_preference_Bachelors         983 non-null    int64
 17  tenancy_preference_Bachelors/Family  983 non-null    int64
```

```
 18  tenancy_preference_Family           983 non-null    int64
dtypes: float64(3), int64(16)
memory usage: 146.0 KB
```

[17]:
```
#changing months of validation
```

[18]:
```
X_val = pd.get_dummies(X_val, columns = ['advertised_month'], dtype = int)
```

[19]:
```
X_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 686 entries, 0 to 685
Data columns (total 19 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   number_of_bedrooms                    686 non-null    int64
 1   floor_area                            686 non-null    int64
 2   current_level                         686 non-null    float64
 3   total_level                           686 non-null    float64
 4   number_of_bathrooms                   686 non-null    int64
 5   advertised_month                      686 non-null    int64
 6   average_rent_bath&bed                 686 non-null    float64
 7   suburb_Adelaide                       686 non-null    int64
 8   suburb_Brisbane                       686 non-null    int64
 9   suburb_Canberra                       686 non-null    int64
 10  suburb_Melbourne                      686 non-null    int64
 11  suburb_Perth                          686 non-null    int64
 12  suburb_Sydney                         686 non-null    int64
 13  furnished_Furnished                   686 non-null    int64
 14  furnished_Semi-Furnished              686 non-null    int64
 15  furnished_Unfurnished                 686 non-null    int64
 16  tenancy_preference_Bachelors          686 non-null    int64
 17  tenancy_preference_Bachelors/Family   686 non-null    int64
 18  tenancy_preference_Family             686 non-null    int64
dtypes: float64(3), int64(16)
memory usage: 102.0 KB
```

[20]:
```
#engineering feature for testing set
```

[21]:
```
X_test = pd.get_dummies(X_test, columns = ['advertised_month'], dtype = int)
```

2 feature scaling using standardisation

[22]:
```
from sklearn.preprocessing import StandardScaler
```

[23]:
```
#training set scaling.
```

[24]:
```
features_to_scale = ['floor_area', 'current_level', 'total_level',␣
 ↪'average_rent_bath&bed' ]
scaler = StandardScaler()
X_train[features_to_scale] = scaler.fit_transform(X_train[features_to_scale])
```

[25]:
```
X_train
```

[25]:

```
      number_of_bedrooms  floor_area  current_level  total_level  \
0                      2    0.579274      -0.579454    -0.488316
1                      2   -0.112940      -0.375746    -0.362558
2                      2    0.348536      -0.375746    -0.362558
3                      2    0.002429      -0.375746    -0.488316
4                      2   -0.574415      -0.579454    -0.614073
…                    …           …              …            …
3311                   3    0.925380       0.235375    -0.111043
3312                   2    1.156118      -0.172039    -0.488316
3313                   2    0.348536       0.031668    -0.111043
3314                   3    2.655914      -0.375746    -0.236801
3315                   2    0.348536       0.235375    -0.111043

      number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                       2               -0.128926                0
1                       1               -0.673384                0
2                       1               -0.673384                0
3                       1               -0.673384                0
4                       2               -0.128926                0
…                     …                       …                …
3311                    2                0.213783                0
3312                    2               -0.128926                0
3313                    2               -0.128926                0
3314                    3                1.308400                0
3315                    2               -0.128926                0

      suburb_Brisbane  suburb_Canberra  suburb_Melbourne  …  suburb_Sydney  \
0                   0                1                 0  …              0
1                   0                1                 0  …              0
2                   0                1                 0  …              0
3                   0                1                 0  …              0
4                   0                1                 0  …              0
…                 …                …                 …  …            …
3311                0                0                 0  …              0
3312                0                0                 0  …              0
3313                0                0                 0  …              0
3314                0                0                 0  …              0
3315                0                0                 0  …              0

      furnished_Furnished  furnished_Semi-Furnished  furnished_Unfurnished  \
```

```
0                       0                       0                       1
1                       0                       1                       0
2                       0                       1                       0
3                       0                       0                       1
4                       0                       0                       1
…                     …                       …                     …
3311                    1                       0                       0
3312                    0                       0                       1
3313                    0                       1                       0
3314                    0                       1                       0
3315                    0                       0                       1

      tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                                 0                                   1
1                                 0                                   1
2                                 0                                   1
3                                 1                                   0
4                                 0                                   1
…                               …                                   …
3311                              1                                   0
3312                              0                                   1
3313                              0                                   1
3314                              0                                   1
3315                              1                                   0

      tenancy_preference_Family  advertised_month_4  advertised_month_5  \
0                             0                   0                   1
1                             0                   0                   1
2                             0                   0                   1
3                             0                   0                   1
4                             0                   1                   0
…                           …                   …                   …
3311                          0                   0                   0
3312                          0                   0                   0
3313                          0                   0                   1
3314                          0                   0                   1
3315                          0                   0                   1

      advertised_month_6
0                       0
1                       0
2                       0
3                       0
4                       0
…                     …
3311                    1
3312                    1
```

```
3313                    0
3314                    0
3315                    0

[3316 rows x 21 columns]
```

[26]:
```
#scaling for validation
```

[27]:
```
feature_to_scale = ['floor_area', 'current_level', 'total_level',␣
 ↪'average_rent_bath&bed' ]
scaler = StandardScaler()
X_val[features_to_scale] = scaler.fit_transform(X_val[features_to_scale])
```

[28]:
```
X_val
```

[28]:
```
     number_of_bedrooms  floor_area  current_level  total_level  \
0                     2   -0.673332      -0.590103    -0.607479
1                     2   -0.251950      -0.685682     2.770869
2                     3    0.191611      -0.590103    -0.374490
3                     1   -0.806400      -0.207785    -0.490985
4                     2   -0.584620      -0.207785    -0.374490
..                  …         …             …          …
978                   2    0.302501       0.174532    -0.141500
979                   3    0.524281      -0.207785    -0.141500
980                   3   -1.440692      -0.207785    -0.490985
981                   1   -0.806400      -0.590103    -0.607479
982                   3    1.411402      -0.781261    -0.490985

     number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                      2              -0.141830                0
1                      2              -0.141830                0
2                      2               0.171511                1
3                      1              -0.596271                0
4                      2              -0.141830                0
..                   …                    …              …
978                    2              -0.141830                0
979                    3               1.075625                0
980                    4               3.935741                0
981                    1              -0.596271                0
982                    3               1.075625                0

     suburb_Brisbane  suburb_Canberra  suburb_Melbourne  …  suburb_Sydney  \
0                  0                0                 1  …              0
1                  0                0                 0  …              1
2                  0                0                 0  …              0
3                  0                0                 0  …              1
4                  1                0                 0  …              0
```

```
..                ...               ...                 ...  ...             ...
978                 0                 0                   0  ...               0
979                 0                 0                   0  ...               0
980                 0                 0                   0  ...               0
981                 0                 0                   0  ...               0
982                 0                 0                   0  ...               0

     furnished_Furnished  furnished_Semi-Furnished  furnished_Unfurnished  \
0                      0                         1                      0
1                      0                         0                      1
2                      0                         0                      1
3                      0                         1                      0
4                      0                         1                      0
..                   ...                       ...                    ...
978                    1                         0                      0
979                    0                         1                      0
980                    1                         0                      0
981                    1                         0                      0
982                    0                         1                      0

     tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                               0                                    0
1                               0                                    1
2                               0                                    1
3                               1                                    0
4                               0                                    1
..                            ...                                  ...
978                             0                                    1
979                             0                                    1
980                             1                                    0
981                             0                                    1
982                             0                                    0

     tenancy_preference_Family  advertised_month_4  advertised_month_5  \
0                            1                   0                   0
1                            0                   0                   0
2                            0                   1                   0
3                            0                   0                   1
4                            0                   1                   0
..                         ...                 ...                 ...
978                          0                   0                   0
979                          0                   0                   0
980                          0                   0                   0
981                          0                   0                   0
982                          1                   0                   0

     advertised_month_6
```

```
0                     1
1                     1
2                     0
3                     0
4                     0
..                  ...
978                   1
979                   1
980                   1
981                   1
982                   1

[983 rows x 21 columns]
```

[29]:
```
#scaling for testing
```

[30]:
```
featuree_to_scale = ['floor_area', 'current_level', 'total_level',␣
 ↪'average_rent_bath&bed' ]
scaler = StandardScaler()
X_test[features_to_scale] = scaler.fit_transform(X_test[features_to_scale])
```

[31]:
```
X_test
```

[31]:
```
     number_of_bedrooms  floor_area  current_level  total_level  \
0                     2   -0.609120      -0.539867    -0.548542
1                     2   -0.293553      -0.632686     2.606330
2                     3    0.038621      -0.539867    -0.330965
3                     1   -0.708772      -0.168590    -0.439754
4                     2   -0.542685      -0.168590    -0.330965
..                  ...         ...            ...          ...
681                   1   -0.874859      -0.168590    -0.330965
682                   3    1.450364      -0.354228    -0.330965
683                   2    0.453840       0.017048    -0.113388
684                   1   -0.728702      -0.632686     1.083288
685                   1   -0.708772      -0.354228    -0.330965

     number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                      2              -0.184552                0
1                      2              -0.184552                0
2                      2               0.099503                1
3                      1              -0.457401                0
4                      2              -0.184552                0
..                   ...                    ...              ...
681                    1              -0.457401                0
682                    3               0.562246                0
683                    2              -0.184552                0
684                    2              -0.034366                0
```

```
685                    1              -0.457401                0

     suburb_Brisbane  suburb_Canberra  suburb_Melbourne  …  suburb_Sydney  \
0                  0                0                 1  …              0
1                  0                0                 0  …              1
2                  0                0                 0  …              0
3                  0                0                 0  …              1
4                  1                0                 0  …              0
..               …                …                 …  …            …
681                1                0                 0  …              0
682                0                0                 0  …              0
683                0                0                 0  …              0
684                0                0                 0  …              1
685                0                0                 0  …              0

     furnished_Furnished  furnished_Semi-Furnished  furnished_Unfurnished  \
0                      0                         1                      0
1                      0                         0                      1
2                      0                         0                      1
3                      0                         1                      0
4                      0                         1                      0
..                   …                         …                    …
681                    0                         0                      1
682                    1                         0                      0
683                    0                         0                      1
684                    0                         1                      0
685                    0                         0                      1

     tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                               0                                    0
1                               0                                    1
2                               0                                    1
3                               1                                    0
4                               0                                    1
..                            …                                    …
681                             0                                    1
682                             0                                    1
683                             0                                    1
684                             0                                    1
685                             0                                    1

     tenancy_preference_Family  advertised_month_4  advertised_month_5  \
0                            1                   0                   0
1                            0                   0                   0
2                            0                   1                   0
3                            0                   0                   1
4                            0                   1                   0
```

```
..                          …                …                …
681                         0                0                1
682                         0                0                1
683                         0                0                0
684                         0                0                1
685                         0                1                0

     advertised_month_6
0                     1
1                     1
2                     0
3                     0
4                     0
..                  …
681                   0
682                   0
683                   1
684                   0
685                   0

[686 rows x 21 columns]
```

2.1 D. Feature Selection

[32]:
```
# <Student to fill this section>

features_list = [
    'number_of_bedrooms', 'floor_area', 'current_level', 'total_level',
    'number_of_bathrooms', 'advertised_month_4', 'average_rent_bath&bed',
    'suburb_Adelaide', 'suburb_Brisbane', 'suburb_Canberra',
    'suburb_Melbourne', 'suburb_Perth', 'suburb_Sydney',
    'furnished_Furnished', 'furnished_Semi-Furnished', 'furnished_Unfurnished',
    'tenancy_preference_Bachelors', 'tenancy_preference_Bachelors/Family',
    'tenancy_preference_Family', 'advertised_month_5', 'advertised_month_6'
]
```

[33]:
```
# @title Feature Selection Explanation

wgt_feat_selection_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Feature Selection Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
```

```
)
wgt_feat_selection_explanation
```

[33]:
```
Textarea(value='', description='Feature Selection Explanation:',
layout=Layout(height='100%', width='auto'), p…
```

2.2 E. Train Machine Learning Model

2.2.1 E.1 Import Algorithm

Provide some explanations on why you believe this algorithm is a good fit

[34]:
```
# <Student to fill this section>
```

[35]:
```
from sklearn.linear_model import ElasticNet
from sklearn.metrics import mean_squared_error as mse
```

3 MODEL 1: hyperparameters —> alpha - 0.1 , l1_ratio - 0.5

[36]:
```
model1 = ElasticNet(alpha = 0.1, l1_ratio = 0.5)
```

[37]:
```
model1.fit(X_train, y_train)
```

[37]:
```
ElasticNet(alpha=0.1)
```

[38]:
```
y_val_pred = model1.predict(X_val)
```

[39]:
```
mse_val = mse(y_val_pred, y_val)
rmse_val = np.sqrt(mse_val)
print("the rmse score on validation is :" , round(rmse_val,2) )
```

```
the rmse score on validation is : 27.5
```

[40]:
```
y_test_pred = model1.predict(X_test)
```

[41]:
```
mse_test = mse(y_test_pred, y_test)
rmse_test = np.sqrt(mse_test)
print("the rmse score for test set is :", round(rmse_test,2))
```

```
the rmse score for test set is : 40.65
```

4 MODEL 2: hyperparameters —> alpha - 0.2 , l1_ratio - 0.6

[42]:
```
model2 = ElasticNet(alpha = 0.2, l1_ratio = 0.6)
```

[43]:
```
model2.fit(X_train, y_train)
```

[43]:
```
ElasticNet(alpha=0.2, l1_ratio=0.6)
```

[44]:
```
y_val_pred1 = model2.predict(X_val)
```

[45]:
```
mse_val1 = mse(y_val_pred1, y_val)
rmse_val1 = np.sqrt(mse_val1)
print("the rmse score on validation set is : ", round(rmse_val1,2))
```

```
the rmse score on validation set is :  28.04
```

[46]:
```
y_test_pred1 = model2.predict(X_test)
```

[47]:
```
mse_test1 = mse(y_test_pred1, y_test)
rmse_test1 = np.sqrt(mse_test1)
print("the rmse score on test set is : ", round(rmse_test1,2))
```

```
the rmse score on test set is :  41.36
```

5 MODEL 3: hyperparameters —> alpha - 0.3 , l1_ratio - 0.7

[48]:
```
model3 = ElasticNet(alpha = 0.3 , l1_ratio = 0.7)
```

[49]:
```
model3.fit(X_train, y_train)
```

[49]:
```
ElasticNet(alpha=0.3, l1_ratio=0.7)
```

[50]:
```
y_val_pred2 = model3.predict(X_val)
```

[51]:
```
mse_val2 = mse(y_val_pred2, y_val)
rmse_val2 = np.sqrt(mse_val2)
print("the rmse score on validation set is : ", round(rmse_val2,2))
```

```
the rmse score on validation set is :  28.27
```

[52]:
```
y_test_pred2 = model3.predict(X_test)
```

[53]:
```
mse_test2 = mse(y_test_pred2, y_test)
rmse_test2 = np.sqrt(mse_test2)
print("the rmse score on test set is : ", round(rmse_test2,2))
```

```
the rmse score on test set is :  41.67
```

6 MODEL 4: hyperparameters —> alpha - 0.09 , l1_ratio - 0.45

[54]:
```
model4 = ElasticNet(alpha = 0.09 , l1_ratio = 0.45)
```

[55]:
```
model4.fit(X_train, y_train)
```

[55]:
```
ElasticNet(alpha=0.09, l1_ratio=0.45)
```

[56]:
```
y_val_pred3 = model4.predict(X_val)
```

[57]:
```
mse_val3 = mse(y_val_pred3, y_val)
rmse_val3 = np.sqrt(mse_val3)
print("the rmse score on validation set is : ", round(rmse_val3,2))
```

```
the rmse score on validation set is :  27.48
```

[58]:
```
y_test_pred3 = model4.predict(X_test)
```

[59]:
```
mse_test3 = mse(y_test_pred3, y_test)
rmse_test3 = np.sqrt(mse_test3)
print("the rmse score on Test set is : ", round(rmse_test3,2))
```

```
the rmse score on Test set is :  40.63
```

7 MODEL 5: hyperparameters —> alpha - 0.08 , l1_ratio - 0.40

[60]:
```
model5 = ElasticNet(alpha = 0.08 , l1_ratio = 0.40)
```

[61]:
```
model5.fit(X_train, y_train)
```

[61]:
```
ElasticNet(alpha=0.08, l1_ratio=0.4)
```

[62]:
```
y_val_pred4 = model5.predict(X_val)
```

[63]:
```
mse_val4 = mse(y_val_pred4, y_val)
rmse_val4 = np.sqrt(mse_val4)
print("the rmse score on validation set is : ", round(rmse_val4,2))
```

```
the rmse score on validation set is :  27.44
```

[64]:
```
y_test_pred4 = model5.predict(X_test)
```

[65]:
```
mse_test4 = mse(y_test_pred4, y_test)
rmse_test4 = np.sqrt(mse_test4)
print("the rmse score on Test set is : ", round(rmse_test4,2))
```

```
the rmse score on Test set is :  40.58
```

8 MODEL 6: hyperparameters —> alpha - 0.07 , l1_ratio - 0.35

[66]:
```
model6 = ElasticNet(alpha = 0.07 , l1_ratio = 0.35)
```

[67]:
```
model6.fit(X_train, y_train)
```

[67]:
```
ElasticNet(alpha=0.07, l1_ratio=0.35)
```

[68]:
```
y_val_pred5 = model6.predict(X_val)
```

[69]:
```
mse_val5 = mse(y_val_pred5, y_val)
rmse_val5 = np.sqrt(mse_val5)
print("the rmse score on validation set is : ", round(rmse_val5,2))
```

```
the rmse score on validation set is :  27.39
```

[70]:
```
y_test_pred5 = model6.predict(X_test)
```

[71]:
```
mse_test5 = mse(y_test_pred5, y_test)
rmse_test5 = np.sqrt(mse_test5)
print("the rmse score on Test set is : ", round(rmse_test5,2))
```

```
the rmse score on Test set is :  40.51
```

9 MODEL 7: hyperparameters —> alpha - 0.06 , l1_ratio - 0.30

[72]:
```
model7 = ElasticNet(alpha = 0.06 , l1_ratio = 0.30)
```

[73]:
```
model7.fit(X_train, y_train)
```

[73]:
```
ElasticNet(alpha=0.06, l1_ratio=0.3)
```

[74]:
```
y_val_pred6 = model7.predict(X_val)
```

[75]:
```
mse_val6 = mse(y_val_pred6, y_val)
rmse_val6 = np.sqrt(mse_val6)
print("the rmse score on validation set is : ", round(rmse_val6,2))
```

```
the rmse score on validation set is :  27.32
```

[76]:
```
y_test_pred6 = model7.predict(X_test)
```

[77]:
```
mse_test6 = mse(y_test_pred6, y_test)
rmse_test6 = np.sqrt(mse_test6)
print("the rmse score on Test set is : ", round(rmse_test6,2))
```

```
the rmse score on Test set is :  40.42
```

10 MODEL 8: hyperparameters —> alpha - 0.05 , l1_ratio - 0.25

[78]:
```
model8 = ElasticNet(alpha = 0.05 , l1_ratio = 0.25)
```

[79]:
```
model8.fit(X_train, y_train)
```

[79]:
```
ElasticNet(alpha=0.05, l1_ratio=0.25)
```

[80]:
```
y_val_pred7 = model8.predict(X_val)
```

[81]:
```
mse_val7 = mse(y_val_pred7, y_val)
rmse_val7 = np.sqrt(mse_val7)
print("the rmse score on validation set is : ", round(rmse_val7,2))
```

```
the rmse score on validation set is :  27.24
```

[82]:
```
y_test_pred7 = model8.predict(X_test)
```

[83]:
```
mse_test7 = mse(y_test_pred7, y_test)
rmse_test7 = np.sqrt(mse_test7)
print("the rmse score on Test set is : ", round(rmse_test7,2))
```

```
the rmse score on Test set is :  40.3
```

11 MODEL 9: hyperparameters —> alpha - 0.04 , l1_ratio - 0.20

[84]:
```
model9 = ElasticNet(alpha = 0.04 , l1_ratio = 0.20)
```

[85]:
```
model9.fit(X_train, y_train)
```

[85]:
```
ElasticNet(alpha=0.04, l1_ratio=0.2)
```

[86]:
```
y_val_pred8 = model9.predict(X_val)
```

[87]:
```
mse_val8 = mse(y_val_pred8, y_val)
rmse_val8 = np.sqrt(mse_val8)
print("the rmse score on validation set is : ", round(rmse_val8,2))
```

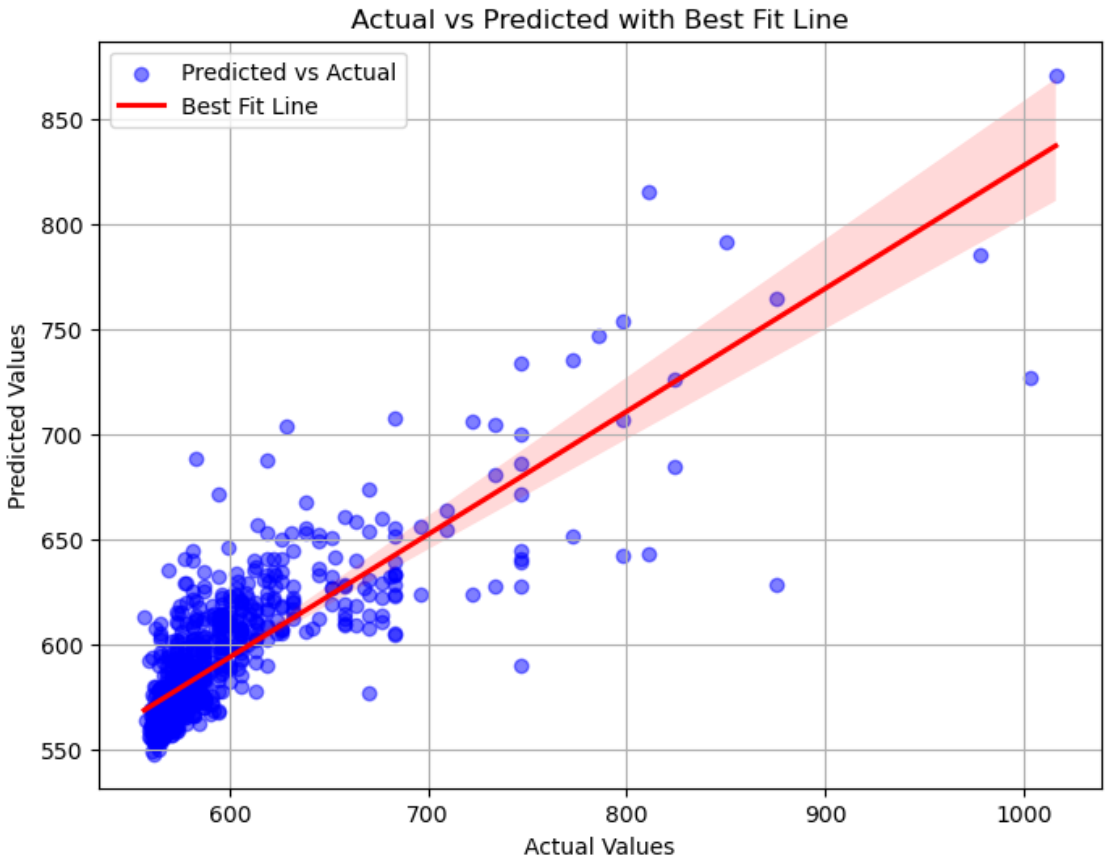
```
the rmse score on validation set is :  27.14
```

[183]:
```
plt.figure(figsize=(8, 6))
plt.scatter(y_val, y_val_pred8, color='blue', alpha=0.5, label='Predicted vs␣
↪Actual')

# Best-fit line (using Seaborn's regression plot)
sns.regplot(x=y_val, y=y_val_pred8, scatter=False, color='red', label='Best Fit␣
↪Line')
```

```
# Labels and title
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted with Best Fit Line')
plt.legend()
plt.grid(True)

plt.show()
```



[ ]:

[89]:
```
y_test_pred8 = model9.predict(X_test)
```

[90]:
```
mse_test8 = mse(y_test_pred8, y_test)
rmse_test8 = np.sqrt(mse_test8)
print("the rmse score on Test set is : ", round(rmse_test8,2))
```

```
the rmse score on Test set is :  40.17
```

[ ]:

[ ]:

12 MODEL 10: hyperparameters —> alpha - 0.01 , l1_ratio - 0.10

[248]:
```
model10 = ElasticNet(alpha = 0.01 , l1_ratio = 0.10)
```

[249]:
```
model10.fit(X_train, y_train)
```

[249]:
```
ElasticNet(alpha=0.01, l1_ratio=0.1)
```

[93]:
```
y_val_pred9 = model10.predict(X_val)
```

[94]:
```
mse_val9 = mse(y_val_pred9, y_val)
rmse_val9 = np.sqrt(mse_val9)
print("the rmse score on validation set is : ", round(rmse_val9,2))
```

```
the rmse score on validation set is :  26.77
```

[95]:
```
y_test_pred9 = model10.predict(X_test)
```

[96]:
```
mse_test9 = mse(y_test_pred9, y_test)
rmse_test9 = np.sqrt(mse_test9)
print("the rmse score on Test set is : ", round(rmse_test9,2))
```

```
the rmse score on Test set is :  39.59
```

13 we will now stop with the experimentation because the returns are fairly diminishing and it isn't changing the prediction by much.

[ ]:

14 WE will now perform the test on future month 7 with the best model and see if it generalises well.

[156]:
```
month_7_val = pd.read_csv('/Users/ratikpant/Desktop/machine learning/␣
 ↪month_07_val')
month_7_test = pd.read_csv('/Users/ratikpant/Desktop/machine learning/␣
 ↪month_07_test')
```

[98]:
```
month_7_val['avg_bed_bath_rent'] = month_7_val.groupby(['number_of_bedrooms',␣
↪'number_of_bathrooms'])['rent'].transform('mean').round(2)
```

[99]:
```
month_7_val
```

[99]:

	number_of_bedrooms	rent	floor_area	current_level	total_level	\
0	2	568.0	800	1.0	2.0	
1	2	565.0	650	1.0	2.0	
2	2	571.0	650	0.0	1.0	
3	2	562.0	800	0.0	1.0	
4	2	564.0	650	0.0	3.0	
..	…	…	…	…	…	
283	2	565.0	900	0.0	2.0	
284	2	565.0	800	1.0	6.0	
285	1	564.0	650	3.0	3.0	
286	2	568.0	1125	2.0	3.0	
287	2	581.0	1350	8.0	14.0	

	suburb	furnished	tenancy_preference	number_of_bathrooms	\
0	Canberra	Unfurnished	Bachelors/Family	1	
1	Canberra	Unfurnished	Family	1	
2	Canberra	Unfurnished	Family	2	
3	Canberra	Unfurnished	Bachelors/Family	1	
4	Canberra	Semi-Furnished	Bachelors/Family	2	
..	…	…	…	…	
283	Perth	Semi-Furnished	Bachelors	2	
284	Perth	Furnished	Bachelors/Family	2	
285	Perth	Semi-Furnished	Bachelors/Family	1	
286	Perth	Unfurnished	Bachelors	2	
287	Perth	Semi-Furnished	Bachelors	2	

	advertised_month	avg_bed_bath_rent
0	7	569.52
1	7	569.52
2	7	580.97
3	7	569.52
4	7	580.97
..	…	…
283	7	580.97
284	7	580.97
285	7	570.04
286	7	580.97
287	7	580.97

[288 rows x 11 columns]

[100]:
```
scaling_features = ['avg_bed_bath_rent', 'floor_area', 'current_level',␣
↪'total_level']
month_7_val[scaling_features] = scaler.
↪fit_transform(month_7_val[scaling_features])
```

[101]:
```
month_7_val = pd.get_dummies(month_7_val, columns = ['suburb', 'furnished',␣
↪'tenancy_preference'], dtype =int)
```

[102]:
```
month_7_val.rename(columns = {'avg_bed_bath_rent' : 'average_rent_bath&bed' })
```

[102]:

	number_of_bedrooms	rent	floor_area	current_level	total_level	\
0	2	568.0	-0.234389	-0.352138	-0.463250	
1	2	565.0	-0.574174	-0.352138	-0.463250	
2	2	571.0	-0.574174	-0.592176	-0.623799	
3	2	562.0	-0.234389	-0.592176	-0.623799	
4	2	564.0	-0.574174	-0.592176	-0.302701	
..	…	…	…	…	…	
283	2	565.0	-0.007865	-0.592176	-0.463250	
284	2	565.0	-0.234389	-0.352138	0.178945	
285	1	564.0	-0.574174	0.127937	-0.302701	
286	2	568.0	0.501812	-0.112101	-0.302701	
287	2	581.0	1.011490	1.328124	1.463336	

	number_of_bathrooms	advertised_month	average_rent_bath&bed	\
0	1	7	-0.556315	
1	1	7	-0.556315	
2	2	7	-0.225573	
3	1	7	-0.556315	
4	2	7	-0.225573	
..	…	…	…	
283	2	7	-0.225573	
284	2	7	-0.225573	
285	1	7	-0.541295	
286	2	7	-0.225573	
287	2	7	-0.225573	

	suburb_Adelaide	suburb_Brisbane	suburb_Canberra	suburb_Melbourne	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	
..	…	…	…	…	
283	0	0	0	0	
284	0	0	0	0	
285	0	0	0	0	
286	0	0	0	0	

```
287                0                0                0                0

     suburb_Perth  suburb_Sydney  furnished_Furnished  \
0               0              0                    0
1               0              0                    0
2               0              0                    0
3               0              0                    0
4               0              0                    0
..            …              …                    …
283             1              0                    0
284             1              0                    1
285             1              0                    0
286             1              0                    0
287             1              0                    0

     furnished_Semi-Furnished  furnished_Unfurnished  \
0                           0                      1
1                           0                      1
2                           0                      1
3                           0                      1
4                           1                      0
..                        …                      …
283                         1                      0
284                         0                      0
285                         1                      0
286                         0                      1
287                         1                      0

     tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                               0                                    1
1                               0                                    0
2                               0                                    0
3                               0                                    1
4                               0                                    1
..                            …                                    …
283                             1                                    0
284                             0                                    1
285                             0                                    1
286                             1                                    0
287                             1                                    0

     tenancy_preference_Family
0                            0
1                            1
2                            1
3                            0
4                            0
```

```
..                          …
283                         0
284                         0
285                         0
286                         0
287                         0

[288 rows x 20 columns]
```

[ ]:

15 matching the columns to what the model9 was trained on using reindex

[103]:
```
month_7_aligned = month_7_val.reindex(columns = X_train.columns, fill_value = 0)
```

[104]:
```
month_7_aligned['rent'] = month_7_val['rent']
```

[105]:
```
month_7_aligned
```

[105]:
```
     number_of_bedrooms  floor_area  current_level  total_level  \
0                     2   -0.234389      -0.352138    -0.463250
1                     2   -0.574174      -0.352138    -0.463250
2                     2   -0.574174      -0.592176    -0.623799
3                     2   -0.234389      -0.592176    -0.623799
4                     2   -0.574174      -0.592176    -0.302701
..                  …         …              …           …
283                   2   -0.007865      -0.592176    -0.463250
284                   2   -0.234389      -0.352138     0.178945
285                   1   -0.574174       0.127937    -0.302701
286                   2    0.501812      -0.112101    -0.302701
287                   2    1.011490       1.328124     1.463336

     number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                      1                      0                0
1                      1                      0                0
2                      2                      0                0
3                      1                      0                0
4                      2                      0                0
..                   …                      …              …
283                    2                      0                0
284                    2                      0                0
285                    1                      0                0
286                    2                      0                0
287                    2                      0                0
```

```
     suburb_Brisbane  suburb_Canberra  suburb_Melbourne  …  \
0                  0                1                 0  …
1                  0                1                 0  …
2                  0                1                 0  …
3                  0                1                 0  …
4                  0                1                 0  …
..               …                …                 …  …
283                0                0                 0  …
284                0                0                 0  …
285                0                0                 0  …
286                0                0                 0  …
287                0                0                 0  …

     furnished_Furnished  furnished_Semi-Furnished  furnished_Unfurnished  \
0                      0                         0                      1
1                      0                         0                      1
2                      0                         0                      1
3                      0                         0                      1
4                      0                         1                      0
..                   …                         …                      …
283                    0                         1                      0
284                    1                         0                      0
285                    0                         1                      0
286                    0                         0                      1
287                    0                         1                      0

     tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                               0                                    1
1                               0                                    0
2                               0                                    0
3                               0                                    1
4                               0                                    1
..                            …                                    …
283                             1                                    0
284                             0                                    1
285                             0                                    1
286                             1                                    0
287                             1                                    0

     tenancy_preference_Family  advertised_month_4  advertised_month_5  \
0                            0                  0                   0
1                            1                  0                   0
2                            1                  0                   0
3                            0                  0                   0
4                            0                  0                   0
..                         …                  …                   …
283                          0                  0                   0
```

```
284                          0                  0                  0
285                          0                  0                  0
286                          0                  0                  0
287                          0                  0                  0

     advertised_month_6   rent
0                     0  568.0
1                     0  565.0
2                     0  571.0
3                     0  562.0
4                     0  564.0
..                  …    …
283                   0  565.0
284                   0  565.0
285                   0  564.0
286                   0  568.0
287                   0  581.0

[288 rows x 22 columns]
```

16 MODEL 4 WITH ALPHA -0.09 and l1_ratio - 0.45 performs the best with the prediction of month 07

[106]:
```
X_fut = month_7_aligned
```

[119]:
```
y_fut =  X_fut.pop('rent')
```

[242]:
```
y_fut_pred = model4.predict(X_fut)
```

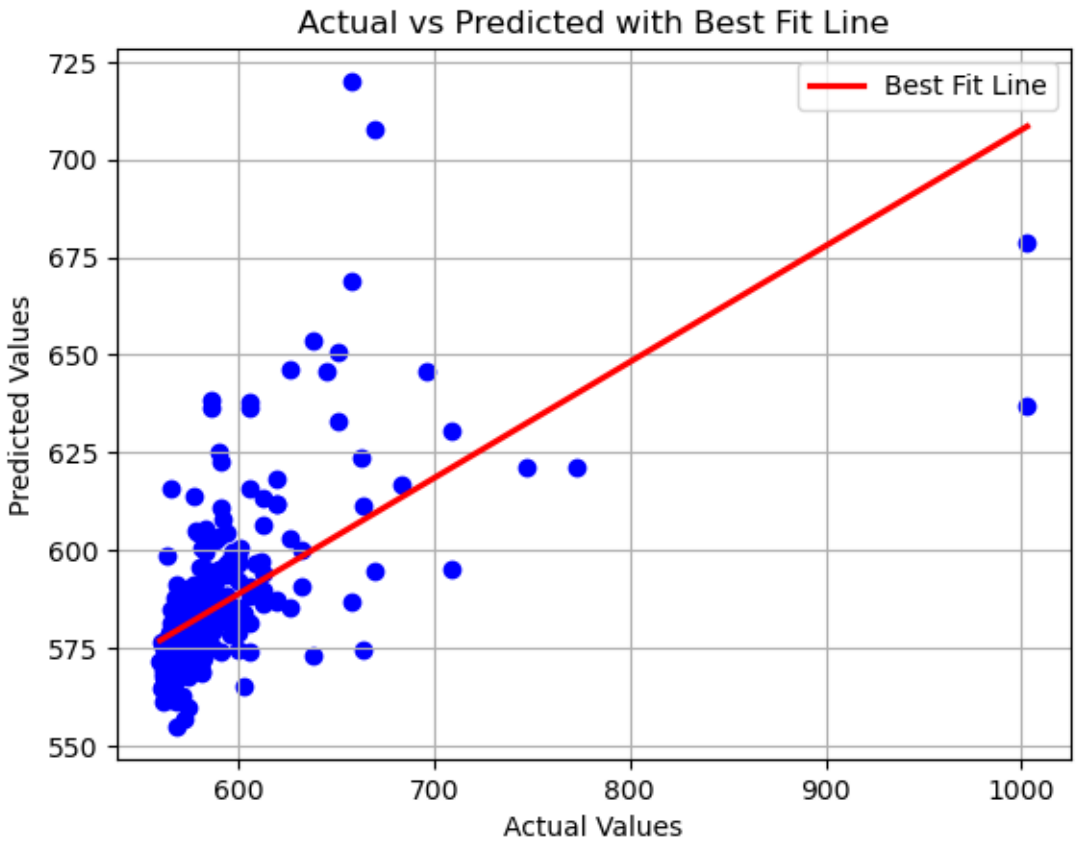
[243]:
```
mse_fut = mse(y_fut_pred,y_fut)
rmse_fut = np.sqrt(mse_fut)
print("the rmse score for validation set for month 7 is : ", round(rmse_fut, 2))
```

```
the rmse score for validation set for month 7 is :  36.39
```

[177]:
```
plt.scatter(y_fut, y_fut_pred,color = 'blue' )
sns.regplot(x=y_fut, y=y_fut_pred, scatter=False , color='red', label='Best Fit␣
 ↪Line', ci=None)

# Labels and title
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted with Best Fit Line')
plt.legend()
plt.grid(True)
```

```
plt.show()
```



17 performance check for month 7 for test set:

[196]:
```
month_7_test
```

[196]:

	number_of_bedrooms	rent	floor_area	current_level	total_level	\
0	2	566.0	720	4.0	4.0	
1	2	587.0	1100	2.0	2.0	
2	3	571.0	800	0.0	1.0	
3	2	564.0	600	0.0	2.0	
4	3	583.0	1150	1.0	2.0	
..	…	…	…	…	…	
673	3	574.0	1500	-1.0	2.0	
674	2	577.0	855	4.0	5.0	
675	2	587.0	1040	2.0	4.0	
676	3	600.0	1750	3.0	5.0	
677	3	613.0	1500	23.0	34.0	

```
       suburb       furnished tenancy_preference  number_of_bathrooms  \
0    Canberra  Semi-Furnished   Bachelors/Family                    2
1    Canberra       Furnished          Bachelors                    2
2    Canberra     Unfurnished   Bachelors/Family                    2
3    Canberra     Unfurnished          Bachelors                    1
4    Canberra     Unfurnished   Bachelors/Family                    2
..        …             …                 …                    …
673     Perth  Semi-Furnished   Bachelors/Family                    3
674     Perth     Unfurnished          Bachelors                    2
675     Perth     Unfurnished          Bachelors                    2
676     Perth  Semi-Furnished   Bachelors/Family                    3
677     Perth  Semi-Furnished             Family                    2

     advertised_month  average_rent_bath&bed
0                   7                 598.70
1                   7                 598.70
2                   7                 601.07
3                   7                 574.33
4                   7                 601.07
..                …                    …
673                 7                 650.79
674                 7                 598.70
675                 7                 598.70
676                 7                 650.79
677                 7                 601.07

[678 rows x 11 columns]
```

[184]:
```
month_7_test['average_rent_bath&bed'] = month_7_test.
↪groupby(['number_of_bedrooms', 'number_of_bathrooms'])['rent'].
↪transform('mean').round(2)
```

[198]:
```
scaling_features = ['average_rent_bath&bed', 'floor_area', 'current_level',␣
↪'total_level']
month_7_test[scaling_features] = scaler.
↪fit_transform(month_7_test[scaling_features])
```

[200]:
```
month_7_test = pd.get_dummies(month_7_test, columns = ['suburb', 'furnished',␣
↪'tenancy_preference'] ,dtype =int)
```

[202]:
```
month_7_test_aligned = month_7_test.reindex(columns = X_train.columns,␣
↪fill_value =0)
```

[203]:
```
month_7_test_aligned
```

[203]:
```
     number_of_bedrooms  floor_area  current_level  total_level  \
0                     2   -0.619113      -0.271703    -0.605285
```

```
1                     2   -0.111422      -0.505552    -0.747953
2                     3   -0.512231      -0.739402    -0.819287
3                     2   -0.779436      -0.739402    -0.747953
4                     3   -0.044621      -0.622477    -0.747953
..                  …         …             …          …
673                   3    0.422989      -0.856326    -0.747953
674                   2   -0.438749      -0.271703    -0.533951
675                   2   -0.191584      -0.505552    -0.605285
676                   3    0.756996      -0.388628    -0.533951
677                   3    0.422989       1.949865     1.534729

     number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                      2              -0.437440                0
1                      2              -0.437440                0
2                      2              -0.398405                0
3                      1              -0.838819                0
4                      2              -0.398405                0
..                   …                    …              …
673                    3               0.420494                0
674                    2              -0.437440                0
675                    2              -0.437440                0
676                    3               0.420494                0
677                    2              -0.398405                0

     suburb_Brisbane  suburb_Canberra  suburb_Melbourne  …  suburb_Sydney  \
0                  0                1                 0  …              0
1                  0                1                 0  …              0
2                  0                1                 0  …              0
3                  0                1                 0  …              0
4                  0                1                 0  …              0
..               …                …                 …  …            …
673                0                0                 0  …              0
674                0                0                 0  …              0
675                0                0                 0  …              0
676                0                0                 0  …              0
677                0                0                 0  …              0

     furnished_Furnished  furnished_Semi-Furnished  furnished_Unfurnished  \
0                      0                         1                      0
1                      1                         0                      0
2                      0                         0                      1
3                      0                         0                      1
4                      0                         0                      1
..                   …                         …                    …
673                    0                         1                      0
674                    0                         0                      1
675                    0                         0                      1
```

```
676                    0                          1                          0
677                    0                          1                          0

     tenancy_preference_Bachelors  tenancy_preference_Bachelors/Family  \
0                               0                                    1
1                               1                                    0
2                               0                                    1
3                               1                                    0
4                               0                                    1
..                            …                                    …
673                             0                                    1
674                             1                                    0
675                             1                                    0
676                             0                                    1
677                             0                                    0

     tenancy_preference_Family  advertised_month_4  advertised_month_5  \
0                            0                   0                   0
1                            0                   0                   0
2                            0                   0                   0
3                            0                   0                   0
4                            0                   0                   0
..                         …                   …                   …
673                          0                   0                   0
674                          0                   0                   0
675                          0                   0                   0
676                          0                   0                   0
677                          1                   0                   0

     advertised_month_6
0                     0
1                     0
2                     0
3                     0
4                     0
..                  …
673                   0
674                   0
675                   0
676                   0
677                   0

[678 rows x 21 columns]
```

[204]:
```
x_future = month_7_test_aligned
```

[205]:
```
x_future['rent'] = month_7_test['rent']
```

[206]:
```
y_future = x_future.pop('rent')
```

[250]:
```
y_future_pred = model10.predict(x_future)
```

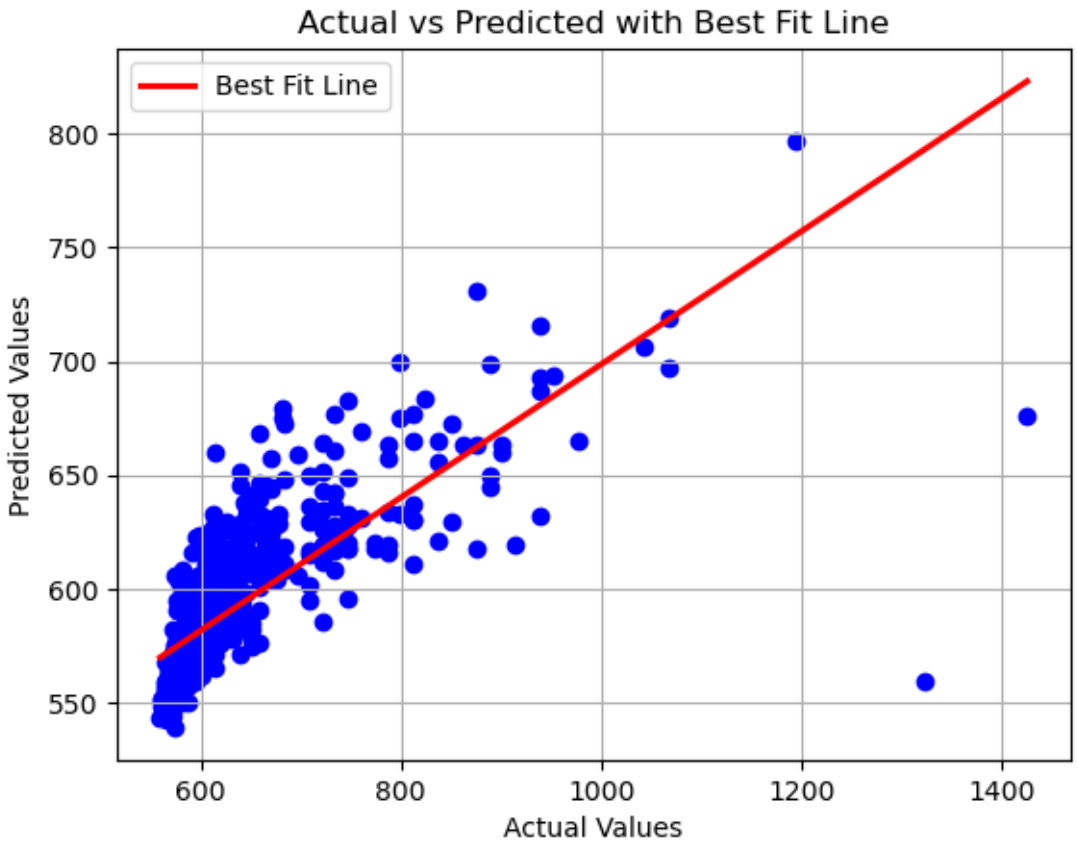
[251]:
```
mse_future = mse(y_future_pred,y_future )
rmse_future = np.sqrt(mse_future)
print(rmse_future)
```

```
76.83099519375777
```

[252]:
```
plt.scatter(y_future, y_future_pred,color = 'blue' )
sns.regplot(x=y_future, y=y_future_pred, scatter=False , color='red',␣
 ↪label='Best Fit Line', ci=None)

# Labels and title
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted with Best Fit Line')
plt.legend()
plt.grid(True)

plt.show()
```



[108]:
```
# @title Algorithm Selection Explanation

wgt_algo_selection_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Algorithm Selection Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_algo_selection_explanation
```

[108]:
```
Textarea(value='', description='Algorithm Selection Explanation:',
layout=Layout(height='100%', width='auto'),…
```

[ ]:
```

```

17.0.1 E.2 Set Hyperparameters

Provide some explanations on why you believe this algorithm is a good fit

[109]:
```
# <Student to fill this section>
```

18 While Model 10 has slightly better RMSE scores, the improvement over Model 9 is marginal (Validation: 0.37, Test: 0.58). Model 9 has simpler hyperparameters (higher alpha and l1_ratio), which means it applies stronger regularization and is less likely to overfit. This makes Model 9 a safer, more robust choice for generalization, especially if the dataset is noisy or small.

[110]:
```
# @title Hyperparameters Selection Explanation

wgt_hyperparams_selection_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Hyperparameters Selection Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_hyperparams_selection_explanation
```

[110]:
```
Textarea(value='', description='Hyperparameters Selection Explanation:',
layout=Layout(height='100%', width='a…
```

18.0.1 E.3 Fit Model

[111]:
```
# <Student to fill this section>
```

18.0.2 E.4 Model Technical Performance

Provide some explanations on model performance

[112]:
```
# <Student to fill this section>
```

[113]:
```
# @title Model Performance Explanation

wgt_model_performance_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Model Performance Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_model_performance_explanation
```

[113]:
```
Textarea(value='', description='Model Performance Explanation:',
layout=Layout(height='100%', width='auto'), p…
```

18.0.3 E.5 Business Impact from Current Model Performance

Provide some analysis on the model impacts from the business point of view

[114]:
```
# <Student to fill this section>
```

[115]:
```
# @title Model Business Impacts Explanation

wgt_model_business_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Model Business Impacts Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_model_business_explanation
```

[115]: Textarea(value='', description='Model Business Impacts Explanation:', layout=Layout(height='100%', width='auto…

18.1 F. Experiment Outcomes

[116]:
```
# @title Experiment Outcomes Explanation

wgt_experiment_outcomes_explanation = widgets.Select(
    options=['Hypothesis Confirmed', 'Hypothesis Partially Confirmed',␣
 ↪'Hypothesis Rejected'],
    value='Hypothesis Rejected',
    description='Experiment Outcomes:',
    disabled=False,
)

wgt_experiment_outcomes_explanation
```

[116]: Select(description='Experiment Outcomes:', index=2, options=('Hypothesis Confirmed', 'Hypothesis Partially Con…

[117]:
```
# @title Experiments Results Explanation

wgt_experiment_results_explanation = widgets.Textarea(
    value=None,
    placeholder='<student to fill this section>',
    description='Experiments Results Explanation:',
    disabled=False,
    style={'description_width': 'initial'},
    layout=widgets.Layout(height="100%", width="auto")
)
wgt_experiment_results_explanation
```

[117]: Textarea(value='', description='Experiments Results Explanation:', layout=Layout(height='100%', width='auto'),…