njy45pw88

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

[ ]:

```
# 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 /usr/local/lib/python3.11/dist-
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
/usr/local/lib/python3.11/dist-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 /usr/local/lib/python3.11/dist-
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 /usr/local/lib/python3.11/dist-
packages (from pandas==2.2.2->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-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 /usr/local/lib/python3.11/dist-
packages (from pandas==2.2.2->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-
packages (from pandas==2.2.2->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 1)) (2025.1)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-
packages (from scikit-learn==1.6.1->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (1.13.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-
packages (from scikit-learn==1.6.1->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 2)) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.11/dist-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 /usr/local/lib/python3.11/dist-packages
(from altair==5.5.0->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (3.1.5)
Requirement already satisfied: jsonschema>=3.0 in
/usr/local/lib/python3.11/dist-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (4.23.0)
```

```
Requirement already satisfied: narwhals>=1.14.2 in
/usr/local/lib/python3.11/dist-packages (from altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (1.25.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-
packages (from altair==5.5.0->-r https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (24.2)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-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>=22.2.0 in /usr/local/lib/python3.11/dist-
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)) (25.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.11/dist-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)) (2024.10.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.11/dist-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)) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-
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)) (0.22.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-
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)) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->altair==5.5.0->-r
https://raw.githubusercontent.com/aso-
uts/labs_datasets/main/36106-mlaa/requirements.txt (line 3)) (3.0.2)

###### Connect to personal Google Drive ######
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).

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

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

1.1.2 0.b Disable Warnings Messages

Do not modify this code before running it

[ ]:

```
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>`

[ ]:

```
# <Student to fill this section>
```

1.1.4 0.d Import Packages

[35]:

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

1.2 A. Project Description

[ ]:

```
# @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])
```

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

[ ]:

```
# @title Experiment ID
```

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

```
BoundedIntText(value=3, description='Experiment ID:', max=3,␣
↪style=DescriptionStyle(description_width='initial…
```

[ ]:

```
# @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
```

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

1.3 B. Experiment Description

[ ]:

```
# @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
```

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

[ ]:

```
# @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
```

```
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

[4]:

```
# 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')
```

[6]:

```
# 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')
```

[5]:

```
# 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')
```

1.5 D. Feature Selection

[ ]:

```
# <Student to fill this section>

features_list = []
```

[ ]:

```
# @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
```

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

1.6 E. Train Machine Learning Model

1.6.1 E.1 Import Algorithm

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

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

```
[13]: from sklearn.neighbors import KNeighborsRegressor
      from sklearn.metrics import mean_squared_error as mse
      from sklearn.preprocessing import StandardScaler
```

2 Training set

```
[11]: X_train
```

```
[11]:       number_of_bedrooms  floor_area  current_level  total_level  \
      0                      2        1100            0.0          2.0
      1                      2         800            1.0          3.0
      2                      2        1000            1.0          3.0
      3                      2         850            1.0          2.0
      4                      2         600            0.0          1.0
      …                    …          …              …          …
      3311                   3        1250            4.0          5.0
      3312                   2        1350            2.0          2.0
      3313                   2        1000            3.0          5.0
      3314                   3        2000            1.0          4.0
      3315                   2        1000            4.0          5.0

            number_of_bathrooms  advertised_month  average_rent_bath&bed  \
      0                       2                 5                 583.42
      1                       1                 5                 569.09
      2                       1                 5                 569.09
      3                       1                 5                 569.09
      4                       2                 4                 583.42
      …                     …                …                     …
```

```
3311                    2                 6                 592.44
3312                    2                 6                 583.42
3313                    2                 5                 583.42
3314                    3                 5                 621.25
3315                    2                 5                 583.42

      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
…               …              …              …               …
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

      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
…            …            …                  …
3311             1              0                    1
3312             1              0                    0
3313             1              0                    0
3314             1              0                    0
3315             1              0                    0

      furnished_Semi-Furnished  furnished_Unfurnished  \
0                            0                      1
1                            1                      0
2                            1                      0
3                            0                      1
4                            0                      1
…                        …                    …
3311                         0                      0
3312                         0                      1
3313                         1                      0
3314                         1                      0
3315                         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
0                             0
1                             0
2                             0
3                             0
4                             0
…                         …
3311                          0
3312                          0
3313                          0
3314                          0
3315                          0

[3316 rows x 19 columns]
```

[17]:
```
features_to_scale = ['current_level', 'total_level', 'floor_area',␣
 ↪'average_rent_bath&bed']
scaler = StandardScaler()
X_train[features_to_scale] = scaler.fit_transform(X_train[features_to_scale] )
```

[19]:
```
X_train = pd.get_dummies(X_train, columns = ['advertised_month'], dtype =int)
```

3 Validation set

[26]:
```
X_val
```

[26]:
```
     number_of_bedrooms  floor_area  current_level  total_level  \
0                     2         560            0.0          1.0
1                     2         750           -0.5         30.0
2                     3         950            0.0          3.0
3                     1         500            2.0          2.0
4                     2         600            2.0          3.0
..                  …         …            …          …
978                   2        1000            4.0          5.0
979                   3        1100            2.0          5.0
```

```
980                   3          214            2.0          2.0
981                   1          500            0.0          1.0
982                   3         1500           -1.0          2.0

     number_of_bathrooms  advertised_month  average_rent_bath&bed  \
0                      2                 6                 585.69
1                      2                 6                 585.69
2                      2                 4                 595.55
3                      1                 5                 571.39
4                      2                 4                 585.69
..                   …                 …                      …
978                    2                 6                 585.69
979                    3                 6                 624.00
980                    4                 6                 714.00
981                    1                 6                 571.39
982                    3                 6                 624.00

     suburb_Adelaide  suburb_Brisbane  suburb_Canberra  suburb_Melbourne  \
0                  0                0                0                 1
1                  0                0                0                 0
2                  1                0                0                 0
3                  0                0                0                 0
4                  0                1                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

     suburb_Perth  suburb_Sydney  furnished_Furnished  \
0               0              0                    0
1               0              1                    0
2               0              0                    0
3               0              1                    0
4               0              0                    0
..            …              …                    …
978             1              0                    1
979             1              0                    0
980             1              0                    1
981             1              0                    1
982             1              0                    0

     furnished_Semi-Furnished  furnished_Unfurnished  \
0                           1                      0
1                           0                      1
2                           0                      1
```

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

[983 rows x 19 columns]
```

[27]:

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

[28]:

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

4 testing set

```
[75]: X_test
```

```
[75]:      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]
```

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

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

```
[ ]:
```

5 model 1 : k = 7, p = 2

```
[191]: model1 = KNeighborsRegressor(n_neighbors = 7 , p=2)
```

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

```
[192]: KNeighborsRegressor(n_neighbors=7)
```

6 validation set

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

```
[194]: mse_val = mse(y_val_pred, y_val)
       rmse_val = np.sqrt(mse_val)
       print("the rmse score is: ", rmse_val)
```

```
the rmse score is:  25.918847998308102
```

7 test set

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

```
[190]: mse_test = mse(y_test_pred, y_test)
       rmse_test = np.sqrt(mse_test)
       print("the rmse score is: ", rmse_test)
```

```
the rmse score is:  36.27935291605646
```

8 model2 k =5 , p=2

[160]:
```
model2 = KNeighborsRegressor(n_neighbors = 5 , p=2)
```

[161]:
```
model2.fit(X_train,y_train)
```

[161]:
```
KNeighborsRegressor()
```

9 val_set

[218]:
```
y_val_pred1 = model2.predict(X_val)
```

[219]:
```
mse_val1 = mse(y_val_pred1, y_val)
rmse_val1 = np.sqrt(mse_val1)
print("the rmse score is: ", rmse_val1)
```

```
the rmse score is:  24.66747753507644
```

10 test_set

[220]:
```
y_test_pred1 = model2.predict(X_test)
```

[221]:
```
mse_test1 = mse(y_test_pred1, y_test)
rmse_test1 = np.sqrt(mse_test1)
print("the rmse score is: ", rmse_test1)
```

```
the rmse score is:  37.85658517875828
```

11 model 3 k = 3 and p =2

[222]:
```
model3 = KNeighborsRegressor(n_neighbors = 3 , p=2)
```

[223]:
```
model3.fit(X_train,y_train)
```

[223]:
```
KNeighborsRegressor(n_neighbors=3)
```

12 val_set

[224]:
```
y_val_pred2 = model3.predict(X_val)
```

[225]:
```
mse_val2 = mse(y_val_pred2, y_val)
rmse_val2 = np.sqrt(mse_val2)
print("the rmse score is: ", rmse_val2)
```

```
the rmse score is:  23.140231651734034
```

13 test_set

[227]:
```python
y_test_pred2 = model3.predict(X_test)
```

[228]:
```python
mse_test2 = mse(y_test_pred2, y_test)
rmse_test2 = np.sqrt(mse_test2)
print("the rmse score is: ", rmse_test2)
```

```
the rmse score is:  39.456529532638406
```

14 model4 k = 5 p =1

[195]:
```python
model4 = KNeighborsRegressor(n_neighbors = 5 , p=1)
```

[196]:
```python
model4.fit(X_train,y_train)
```

[196]:
```
KNeighborsRegressor(p=1)
```

15 val_set

[229]:
```python
y_val_pred3 = model4.predict(X_val)
```

[230]:
```python
mse_val3 = mse(y_val_pred3, y_val)
rmse_val3 = np.sqrt(mse_val3)
print("the rmse score is: ", rmse_val3)
```

```
the rmse score is:  23.02404623868304
```

16 test set

[231]:
```python
y_test_pred3 = model4.predict(X_test)
```

[232]:
```python
mse_test3 = mse(y_test_pred3, y_test)
rmse_test3 = np.sqrt(mse_test3)
print("the rmse score is: ", rmse_test3)
```

```
the rmse score is:  36.78030977532919
```

17 model5. k = 3 p= 1

[233]:
```python
model5 = KNeighborsRegressor(n_neighbors = 3 , p=1)
```

[234]:

```
model5.fit(X_train,y_train)
```

[234]:

```
KNeighborsRegressor(n_neighbors=3, p=1)
```

18 val_test

[235]:

```
y_val_pred4 = model5.predict(X_val)
```

[236]:

```
mse_val4 = mse(y_val_pred4, y_val)
rmse_val4 = np.sqrt(mse_val4)
print("the rmse score is: ", rmse_val4)
```

```
the rmse score is:  21.575746562736075
```

19 test_set

[237]:

```
y_test_pred4 = model5.predict(X_test)
```

[238]:

```
mse_test4 = mse(y_test_pred4, y_test)
rmse_test4 = np.sqrt(mse_test4)
print("the rmse score is: ", rmse_test4)
```

```
the rmse score is:  36.27935291605646
```

20 model 6 k =1 p =1

[239]:

```
model6 = KNeighborsRegressor(n_neighbors = 1 , p=1)
```

[240]:

```
model6.fit(X_train,y_train)
```

[240]:

```
KNeighborsRegressor(n_neighbors=1, p=1)
```

21 val_set

[241]:

```
y_val_pred5 = model6.predict(X_val)
```

[242]:

```
mse_val5 = mse(y_val_pred5, y_val)
rmse_val5 = np.sqrt(mse_val5)
print("the rmse score is: ", rmse_val5)
```

```
the rmse score is:  19.233371902852113
```

22 test_set

[243]:

```
y_test_pred5 = model6.predict(X_test)
```

[244]:

```
mse_test5 = mse(y_test_pred5, y_test)
rmse_test5 = np.sqrt(mse_test5)
print("the rmse score is: ", rmse_test5)
```

```
the rmse score is:  38.30278708582802
```

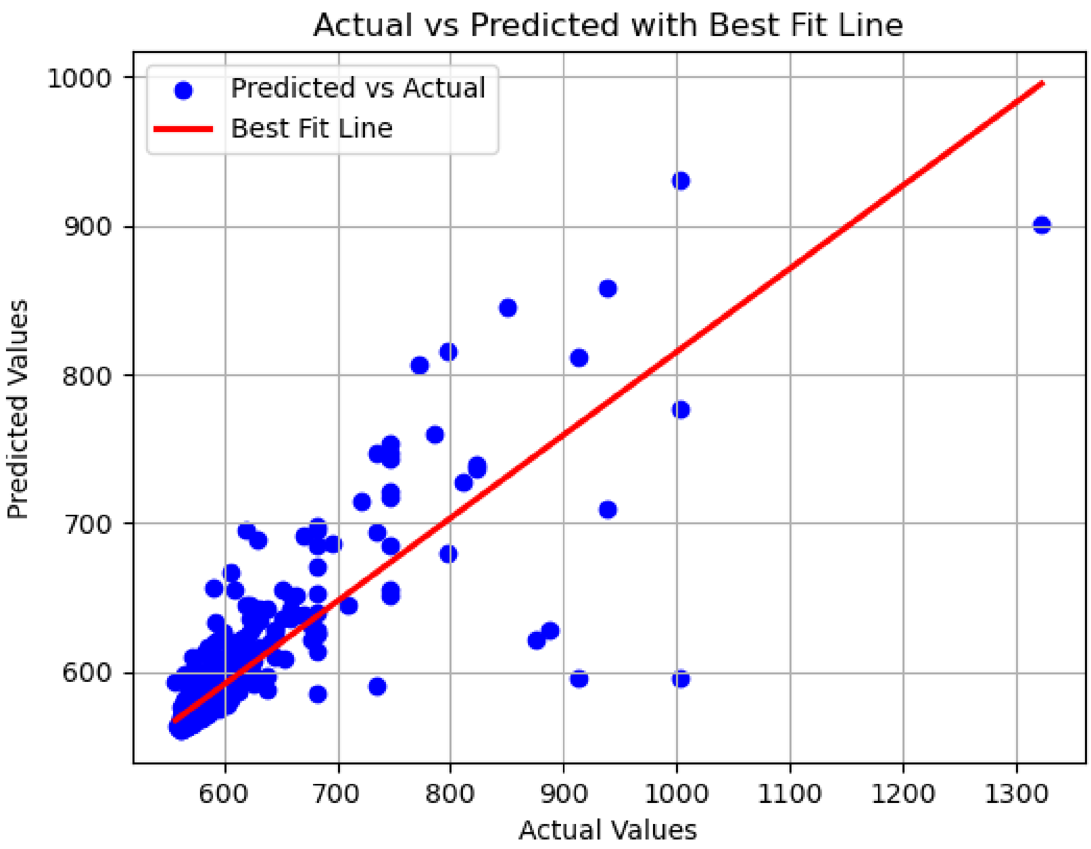
23 Model5: KNN with k=3 and p=1 (Manhattan Distance) works best because it finds a good balance between accuracy and stability. Manhattan Distance handles outliers better than Euclidean, which helps since the dataset has a few outliers. k=3 keeps predictions stable without making them too general. This setup works well for rent prediction because it captures local patterns while avoiding big errors from extreme values.

[249]:

```
plt.scatter(y_test , y_test_pred4, color = 'blue', label = 'Predicted vs␣
 ↪Actual')
sns.regplot(x=y_test, y=y_test_pred4, 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()
```



```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

24 performing it on the future month (7) on validation set and test set. and check how does it performs on it.

```
[245]: 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')
```

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

```
[258]: month_7_val['average_rent_bath&bed'] = month_7_val.
↪groupby(['number_of_bedrooms', 'number_of_bathrooms'])['rent'].
↪transform('mean').round(2)
```

```
[260]: month_7_val_alligned = month_7_val.reindex(columns = X_train.columns ,␣
↪fill_value = 0)
```

```
[262]: month_7_val_alligned['rent'] = month_7_val['rent']
```

```
[263]: month_7_val_alligned
```

```
[263]:      number_of_bedrooms  floor_area  current_level  total_level  \
0                     2         800            1.0          2.0
1                     2         650            1.0          2.0
2                     2         650            0.0          1.0
3                     2         800            0.0          1.0
4                     2         650            0.0          3.0
..                  …         …            …          …
283                   2         900            0.0          2.0
284                   2         800            1.0          6.0
285                   1         650            3.0          3.0
286                   2        1125            2.0          3.0
287                   2        1350            8.0         14.0

     number_of_bathrooms  average_rent_bath&bed  suburb_Adelaide  \
0                      1                 569.52                0
1                      1                 569.52                0
2                      2                 580.97                0
3                      1                 569.52                0
4                      2                 580.97                0
..                   …                    …              …
283                    2                 580.97                0
284                    2                 580.97                0
285                    1                 570.04                0
286                    2                 580.97                0
287                    2                 580.97                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]
```

```
[264]: scaling_features = ['average_rent_bath&bed', 'floor_area', 'current_level',␣
        ↪'total_level']
```

```
[265]: month_7_val_alligned[scaling_features] = scalerr.
        ↪fit_transform(month_7_val_alligned[scaling_features])
```

```
[266]: month_7_val_alligned
```

```
[266]:      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              -0.556315                0
       1                      1              -0.556315                0
       2                      2              -0.225573                0
       3                      1              -0.556315                0
       4                      2              -0.225573                0
       ..                   …                  …              …
       283                    2              -0.225573                0
       284                    2              -0.225573                0
       285                    1              -0.541295                0
```

```
286                    2              -0.225573                 0
287                    2              -0.225573                 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]
```

[267]:

```
X_fut =month_7_val_alligned
```

[268]:

```
y_fut =X_fut.pop('rent')
```

[356]:

```
y_fut_pred = model5.predict(X_fut)
```

[357]:

```
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))
```

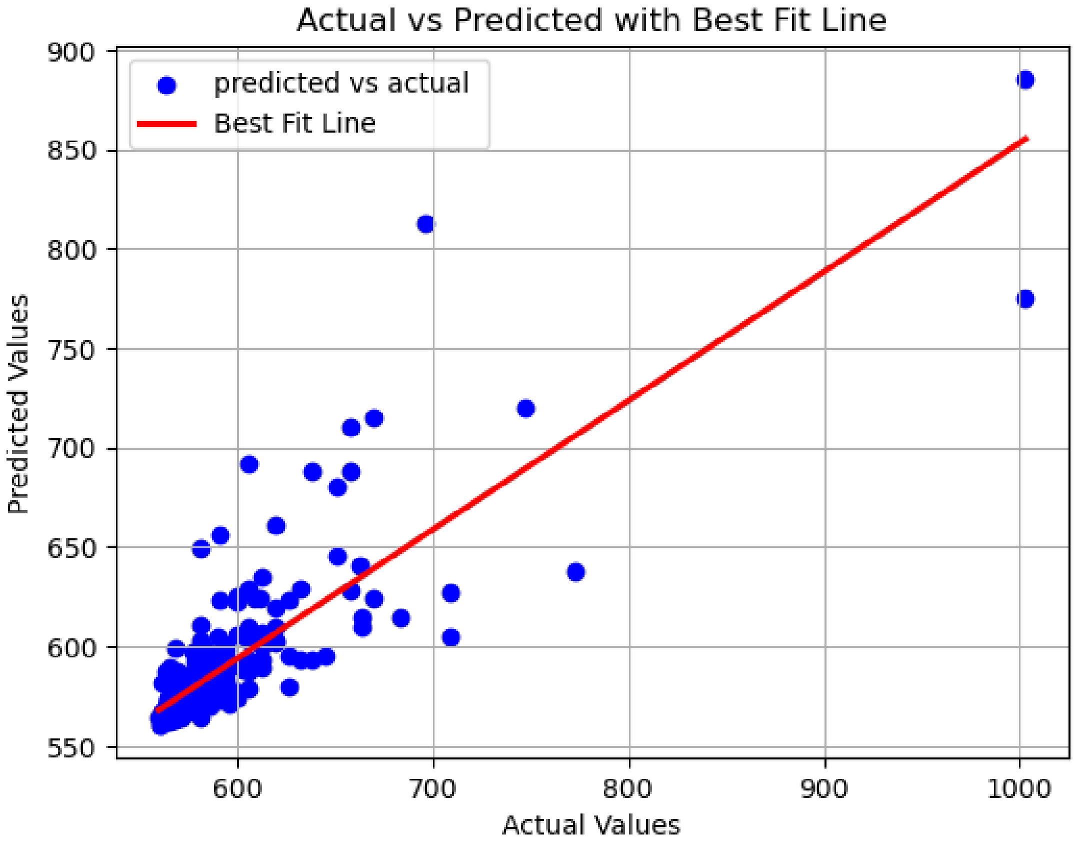
```
the rmse score for validation set for month 7 is :  29.04
```

[349]:

```
plt.scatter(y_fut, y_fut_pred,color = 'blue', label = 'predicted vs actual ' )
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()
```



25 the best performance is coming from model 4 where k =5 and p =1 manhattan distance

26 lets try this on the final test set where all the months are 7

```
[294]: month_7_test
```

```
[294]:      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
0                   7
1                   7
2                   7
3                   7
4                   7
..                …
673                 7
674                 7
675                 7
676                 7
677                 7

[678 rows x 10 columns]
```

[295]:
```
month_7_test = pd.get_dummies(month_7_test, columns = ['suburb',␣
  ↪'furnished','tenancy_preference' ], dtype = int)
```

[296]:
```
month_7_test['average_rent_bath&bed'] = month_7_test.
  ↪groupby(['number_of_bedrooms', 'number_of_bathrooms'])['rent'].
  ↪transform('mean').round(2)
```

[298]:
```
month_7_test_aligned = month_7_test.reindex(columns = X_train.columns,␣
  ↪fill_value =0)
```

[300]:
```
scale_features = ['average_rent_bath&bed', 'floor_area', 'current_level',␣
  ↪'total_level']
```

[301]:
```
month_7_test_aligned[scale_features] = scaler.
  ↪fit_transform(month_7_test_aligned[scale_features])
```

27 completely pre-processed dataset for predicting rent using knn model 4 k =5 and p =1 manhatan distance

[304]:

```
month_7_test_aligned['rent'] = month_7_test ['rent']
```

[305]:

```
x_future = month_7_test_aligned
```

[306]:

```
y_future = x_future.pop('rent')
```

[354]:

```
y_future_pred = model5.predict(x_future)
```

[355]:

```
mse_future = mse(y_future_pred,y_future)
rmse_future = np.sqrt(mse_future)
print("the rmse score for validation set for month 7 is : ", round(rmse_future, 2))
```

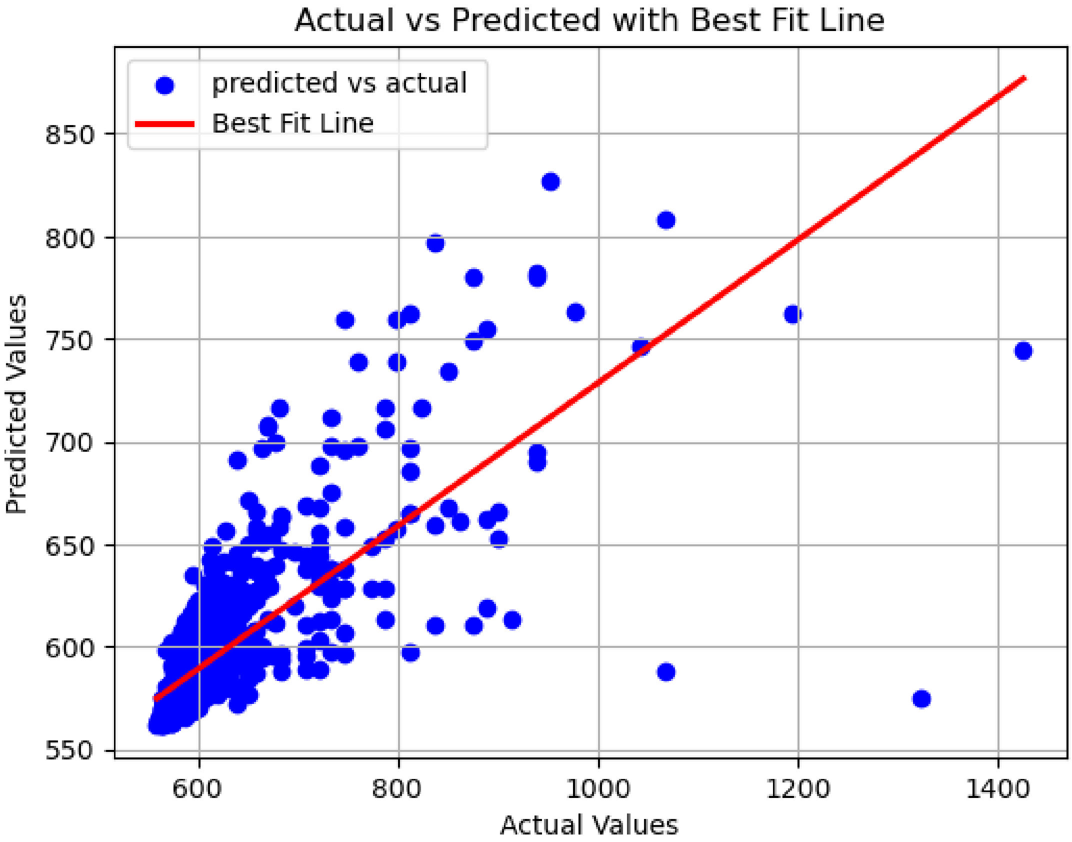
```
the rmse score for validation set for month 7 is :  70.33
```

[347]:

```
plt.scatter(y_future, y_future_pred,color = 'blue', label = 'predicted vs actual ' )
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()
```



28 the best performing knn model on future month (7) & keeping all the previous months as 0, is model number 4::: k =5 and p =1

```
[ ]: # @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
```

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

28.0.1 E.2 Set Hyperparameters

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

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

[ ]:
```
# @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
```

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

28.0.2 E.3 Fit Model

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

28.0.3 E.4 Model Technical Performance

Provide some explanations on model performance

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

[ ]:
```
# @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
```

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

28.0.4 E.5 Business Impact from Current Model Performance

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

[ ]:

```
# <Student to fill this section>
```

[ ]:

```
# @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
```

```
Textarea(value='', description='Model Business Impacts Explanation:',␣
 ↪layout=Layout(height='100%', width='auto…
```

28.1 F. Experiment Outcomes

[ ]:

```
# @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
```

```
Select(description='Experiment Outcomes:', index=2, options=('Hypothesis␣
 ↪Confirmed', 'Hypothesis Partially Con…
```

[ ]:

```
# @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
```

```
Textarea(value='', description='Experiments Results Explanation:',␣
 ↪layout=Layout(height='100%', width='auto'),…
```