

```
!apt-get install -y texlive-xetex texlive-fonts-recommended texlive-
generic-recommended
!pip install nbconvert

Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
E: Unable to locate package texlive-generic-recommended
Requirement already satisfied: nbconvert in
/usr/local/lib/python3.10/dist-packages (6.5.4)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (4.9.3)
Requirement already satisfied: beautifulsoup4 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (4.11.2)
Requirement already satisfied: bleach in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (6.1.0)
Requirement already satisfied: defusedxml in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.7.1)
Requirement already satisfied: entrypoints>=0.2.2 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.4)
Requirement already satisfied: jinja2>=3.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (3.1.2)
Requirement already satisfied: jupyter-core>=4.7 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (5.5.0)
Requirement already satisfied: jupyterlab-pygments in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.2.2)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (2.1.3)
Requirement already satisfied: mistune<2,>=0.8.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.8.4)
Requirement already satisfied: nbclient>=0.5.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.9.0)
Requirement already satisfied: nbformat>=5.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (5.9.2)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (23.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (1.5.0)
Requirement already satisfied: pygments>=2.4.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (2.16.1)
Requirement already satisfied: tinycss2 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (1.2.1)
Requirement already satisfied: traitlets>=5.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (5.7.1)
Requirement already satisfied: platformdirs>=2.5 in
/usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.7-
>nbconvert) (4.0.0)
Requirement already satisfied: jupyter-client>=6.1.12 in
/usr/local/lib/python3.10/dist-packages (from nbclient>=0.5.0-
>nbconvert) (6.1.12)
```

```
Requirement already satisfied: fastjsonschema in
/usr/local/lib/python3.10/dist-packages (from nbformat>=5.1-
>nbconvert) (2.19.0)
Requirement already satisfied: jsonschema>=2.6 in
/usr/local/lib/python3.10/dist-packages (from nbformat>=5.1-
>nbconvert) (4.19.2)
Requirement already satisfied: soupsieve>1.2 in
/usr/local/lib/python3.10/dist-packages (from beautifulsoup4-
>nbconvert) (2.5)
Requirement already satisfied: six>=1.9.0 in
/usr/local/lib/python3.10/dist-packages (from bleach->nbconvert)
(1.16.0)
Requirement already satisfied: webencodings in
/usr/local/lib/python3.10/dist-packages (from bleach->nbconvert)
(0.5.1)
Requirement already satisfied: attrs>=22.2.0 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=5.1->nbconvert) (23.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=5.1->nbconvert) (2023.11.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=5.1->nbconvert) (0.31.0)
Requirement already satisfied: rpds-py>=0.7.1 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=5.1->nbconvert) (0.13.0)
Requirement already satisfied: pyzmq>=13 in
/usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12-
>nbclient>=0.5.0->nbconvert) (23.2.1)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12-
>nbclient>=0.5.0->nbconvert) (2.8.2)
Requirement already satisfied: tornado>=4.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12-
>nbclient>=0.5.0->nbconvert) (6.3.2)
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
import matplotlib.pyplot as plt

# Load data from CSV file
file_path = "/content/COMP1801_CourseworkDataset1_tabular (1).csv"
df = pd.read_csv(file_path)
```

```

# Assuming 'Lifespan' is dependent variable, and other columns are
features
X = df.drop('Lifespan', axis=1)
y = df['Lifespan']

# Split the data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Identify categorical columns
categorical_cols = ['partType', 'microstructure', 'seedLocation',
'castType']

# Create a column transformer
preprocessor = ColumnTransformer(
    transformers=[
        ('num', 'passthrough',
X.columns.difference(categorical_cols)),
        ('cat', OneHotEncoder(), categorical_cols)
    ])

# Create a pipeline with preprocessing and linear regression
model = Pipeline(steps=[('preprocessor', preprocessor),
                        ('regressor', LinearRegression())])

# Train the model
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

# Print evaluation metrics
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

# Plot actual vs. predicted values
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Lifespan")
plt.ylabel("Predicted Lifespan")
plt.title("Actual vs. Predicted Lifespan")
plt.show()

Mean Squared Error: 49179.53422115317
R-squared: 0.8286019504028523

```

Actual vs. Predicted Lifespan

