

BDM 1034 - Application Design for Big Data Lab5

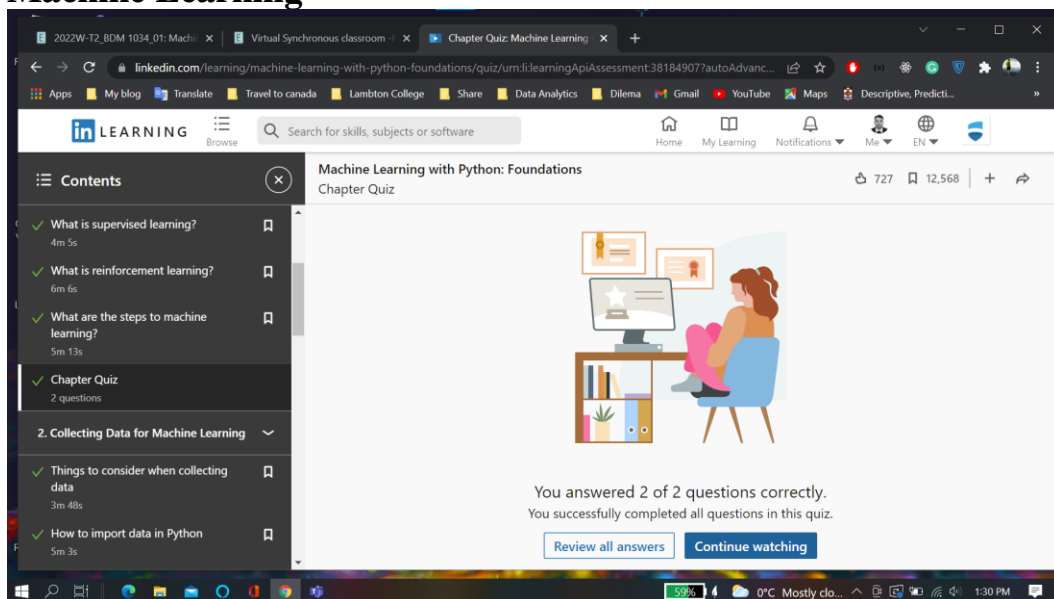
Submitted by: Aadarsha Chapagain

Student ID:C0825975

Submitted to: Prof. Teresa Zhu

Here I have attached the certificate I achieved from LinkedIn learning for Course “**Machine Learning with Python: Foundations**” along with the Screenshot of score I achieved

Chapter 1: Machine Learning



Chapter 2

Collecting Data for Machine Learning

The screenshot shows a Jupyter Notebook interface with a browser window in the background. The notebook is titled "02_02e" and contains a code cell with the following Python code:

```
In [9]: brics6 = pd.read_excel("brics.xlsx", sheet_name = "Summits")
brics6
```

The output of the code is a DataFrame with 13 rows and 6 columns: summit, date, host, leader, and location. The data represents the BRICS summits from 2009 to 2020.

	summit	date	host	leader	location
0	1st	June 16th, 2009	Russia	Dmitry Medvedev	Yekaterinburg (Sevastianov's House)
1	2nd	April 15th, 2010	Brazil	Luiz Inácio Lula da Silva	Brasília (Itamaraty Palace)
2	3rd	April 14th, 2011	China	Hu Jintao	Sanya (Sheraton Sanya Resort)
3	4th	March 29th, 2012	India	Manmohan Singh	New Delhi (Taj Mahal Hotel)
4	5th	March 26th – 27th, 2013	South Africa	Jacob Zuma	Durban (Durban ICC)
5	6th	July 14th – 17th, 2014	Brazil	Dilma Rousseff	Fortaleza (Centro de Eventos do Ceará)
6	7th	July 8th – 9th, 2015	Russia	Vladimir Putin	Ufa (Congress Hall)
7	8th	October 15th – 16th, 2016	India	Narendra Modi	Banaulim (Taj Exotica)
8	9th	September 3th – 5th, 2017	China	Xi Jinping	Xiamen (Xiamen International Conference Center)
9	10th	July 25th – 27th, 2018	South Africa	Cyril Ramaphosa	Johannesburg (Sandton Convention Centre)
10	11th	November 13th – 14th, 2019	Brazil	Jair Bolsonaro	Brasília (Itamaraty Palace)
11	12th	November 17th, 2020	Russia	Vladimir Putin	Saint Petersburg (video conference)
12	13th	TBA	India	Narendra Modi	New Delhi

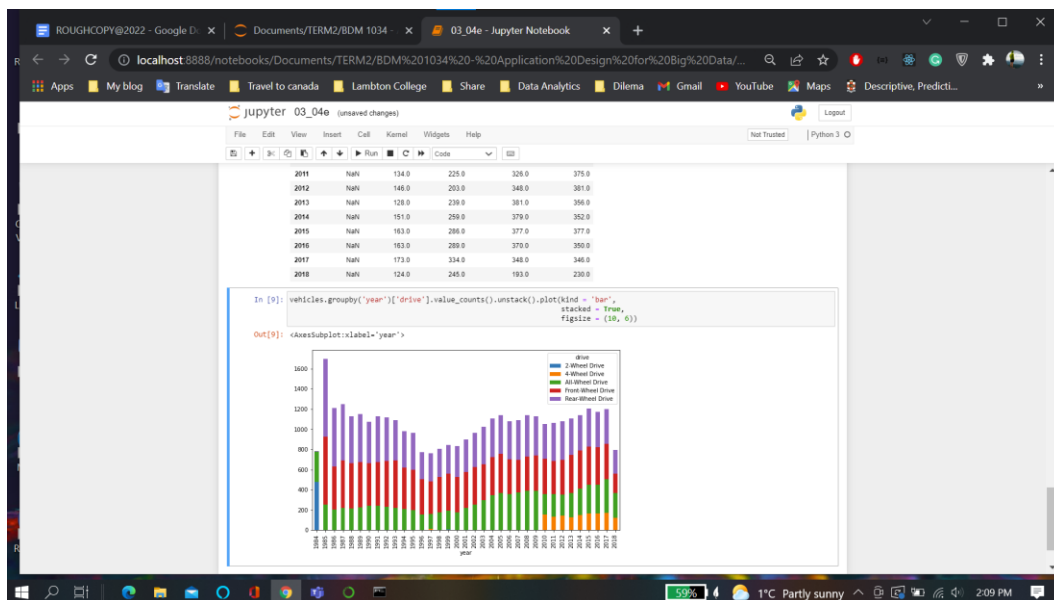
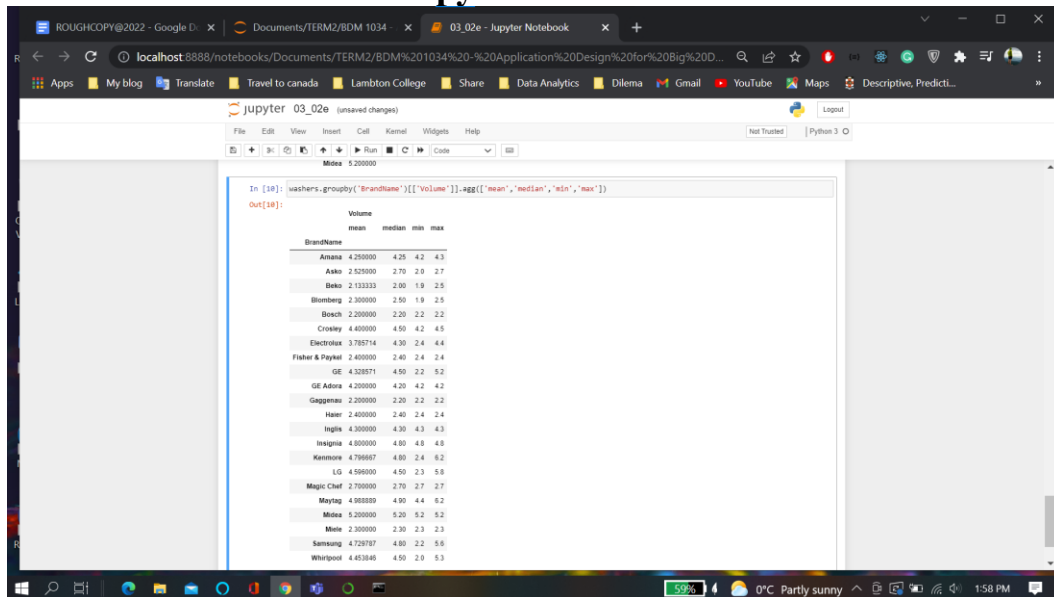
Chapter Quiz:

The screenshot shows a quiz completion screen for "Machine Learning with Python: Foundations". The left sidebar contains a table of contents with the following items:

- Introduction
- 1. Machine Learning
- 2. Collecting Data for Machine Learning
 - Things to consider when collecting data (3m 48s)
 - How to import data in Python (5m 3s)
 - Chapter Quiz (2 questions)
- 3. Understanding Data for Machine Learning
 - Describe your data (3m 33s)
 - How to summarize data in Python (6m)
 - Visualize your data (3m 42s)
 - How to visualize data in Python (6m 23s)

The main content area shows the quiz results for the "Chapter Quiz". It states: "You answered 2 of 2 questions correctly. You successfully completed all questions in this quiz." Below this, there are two buttons: "Review all answers" and "Continue watching".

Chapter 3: How to summarise data in python




Chapter quiz:

Contents

- 3. Understanding Data for Machine Learning
 - Describe your data 3m 33s
 - How to summarize data in Python 6m
 - Visualize your data 3m 42s
 - How to visualize data in Python 6m 23s
 - Chapter Quiz 2 questions
- 4. Preparing Data for Machine Learning
 - Common data quality issues 3m 42s
 - How to resolve missing data in Python 7m 34s
 - Normalizing your data 4m 39s
 - How to normalize data in Python 4m 38s

Machine Learning with Python: Foundations Chapter Quiz

727 12,568



You answered 2 of 2 questions correctly.
You successfully completed all questions in this quiz.

[Review all answers](#) [Continue watching](#)

Chapter 4: Resolve Missing Data in python:

ROUGH COPY@2022 - Google D... Documents/TERM2/BDM 1034 x 04_02e - Jupyter Notebook x +

localhost:8888/notebooks/Documents/TERM2/BDM%201034%20-%20Application%20Design%20for%20Big%20D...
Apps My blog Translate Travel to canada Lambton College Share Data Analytics Dilema Gmail YouTube Maps Descriptive, Predicti...

jupyter 04_02e (autosaved) Logout

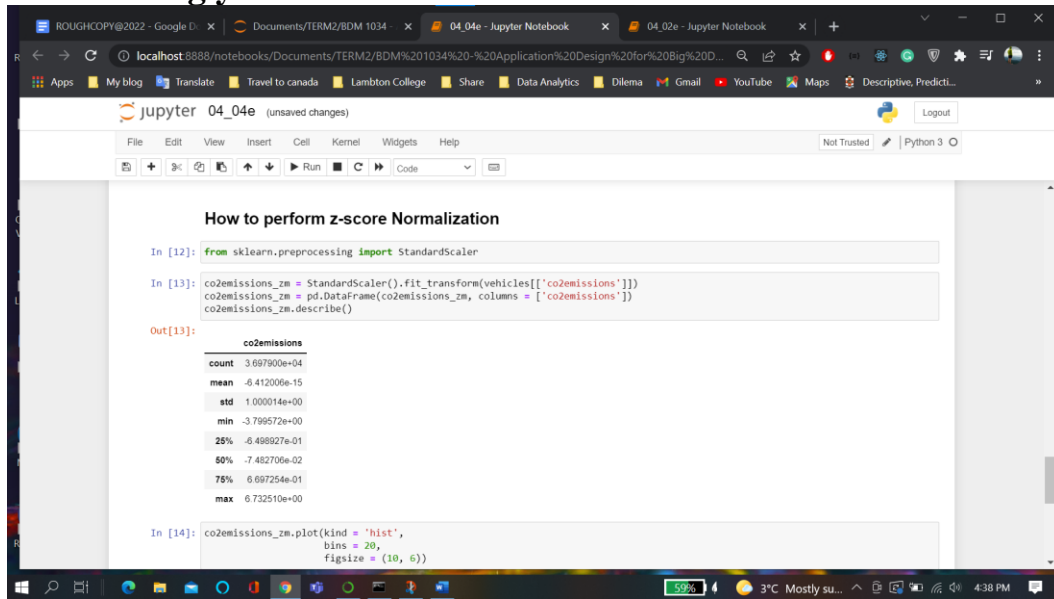
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [13]: mask = (students['City'] == 'Niles') & (students['State'] == 'MI')
students.loc[mask, 'Zip'] = 49120
students
```

Out[13]:

	ID	FirstName	LastName	Major	Age	Gender	City	State	Zip
2	2075	Becca	Swanson	Marketing	22.0	Female	Chicago	IL	60608.0
13	6504	Charlette	Woods	Business Technology	18.0	Female	Chicago	NaN	60608.0
14	6768	Caroline	Marsh	Business Analytics	22.0	Female	Niles	MI	NaN
15	7511	Laila	Carroll	Marketing	20.0	Female	New York	NY	10001.0
16	7965	Rocco	Decola	Finance	21.0	Male	Oakland	CA	94609.0
17	9232	Julie	Holmes	Business Technology	18.0	Female	Webster	NY	14580.0
18	9268	Albert	Palmer	Management	21.0	Male	Detroit	MI	48201.0
19	9941	Zoya	Doyle	Business Analytics	20.0	Female	Chicago	IL	60608.0

Normalizing your Data:



A Jupyter Notebook interface titled '04_04e' showing the process of z-score normalization. The notebook is running on a local host. The code in the first cell imports `StandardScaler` from `sklearn.preprocessing`. The second cell creates a `StandardScaler` object and applies it to the `co2emissions` column of a `vehicles` dataset, storing the result in `co2emissions_zm`. The third cell displays the statistical summary of the original `co2emissions` data. The fourth cell plots a histogram of the normalized data.

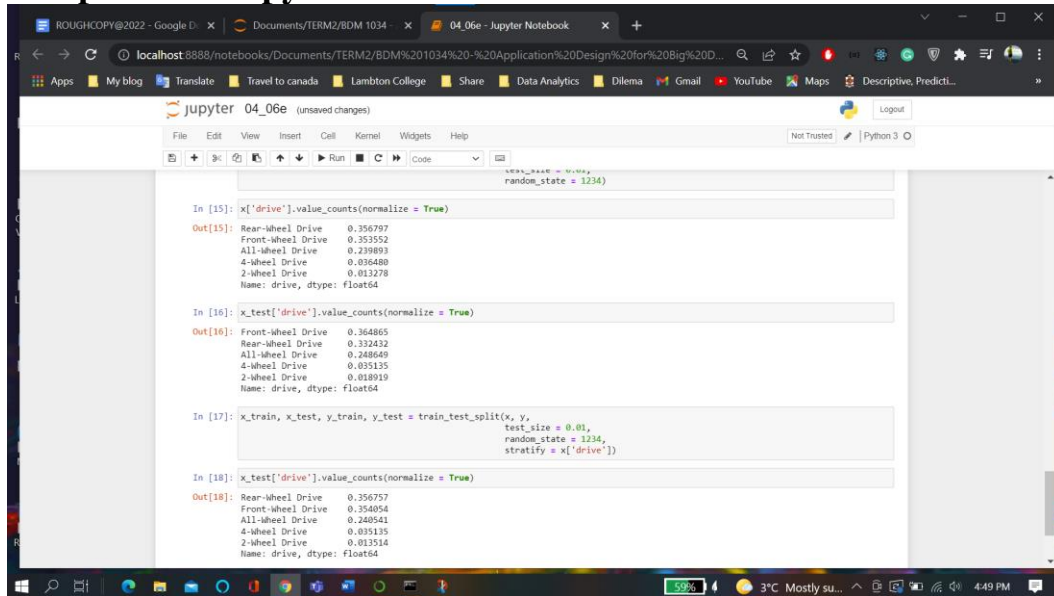
```
In [12]: from sklearn.preprocessing import StandardScaler

In [13]: co2emissions_zm = StandardScaler().fit_transform(vehicles[['co2emissions']])
co2emissions_zm = pd.DataFrame(co2emissions_zm, columns = ['co2emissions'])
co2emissions_zm.describe()

Out[13]:
co2emissions
count    3.697900e+04
mean     -6.412006e-15
std       1.000014e+00
min      -3.799572e+00
25%      -6.488927e-01
50%      -7.482706e-02
75%       6.697254e-01
max       6.732510e+00

In [14]: co2emissions_zm.plot(kind = 'hist',
                               bins = 20,
                               figsize = (10, 6))
```

Sample Data in python



A Jupyter Notebook interface titled '04_06e' showing the generation of sample data and its normalization. The code in the first cell generates a random state. The second cell uses `value_counts` to count the frequency of each drive type in the `x` dataset, with `normalize = True`. The third cell uses `value_counts` to count the frequency of each drive type in the `x_test` dataset, with `normalize = True`. The fourth cell uses `train_test_split` to split the data into training and testing sets, with `test_size = 0.01`, `random_state = 1234`, and `stratify = x['drive']`. The fifth cell uses `value_counts` to count the frequency of each drive type in the `x_test` dataset, with `normalize = True`.

```
In [15]: x['drive'].value_counts(normalize = True)

Out[15]:
Rear-wheel Drive    0.356797
Front-wheel Drive   0.353552
All-wheel Drive     0.239893
4-wheel Drive       0.036480
2-wheel Drive       0.013278
Name: drive, dtype: Float64

In [16]: x_test['drive'].value_counts(normalize = True)

Out[16]:
Front-wheel Drive    0.364865
Rear-wheel Drive     0.332432
All-wheel Drive      0.248649
4-wheel Drive        0.035135
2-wheel Drive        0.018919
Name: drive, dtype: Float64

In [17]: x_train, x_test, y_train, y_test = train_test_split(x, y,
                                                             test_size = 0.01,
                                                             random_state = 1234,
                                                             stratify = x['drive'])

In [18]: x_test['drive'].value_counts(normalize = True)

Out[18]:
Rear-wheel Drive    0.356757
Front-wheel Drive   0.354054
All-wheel Drive     0.240541
4-wheel Drive       0.035135
2-wheel Drive       0.013514
Name: drive, dtype: Float64
```

Chapter Quiz

Contents

✓ Normalizing your data
4m 39s

✓ How to normalize data in Python
4m 38s

✓ Sampling your data
4m 7s

✓ How to sample data in Python
6m 35s

✓ Reducing the dimensionality of your data
3m 24s

✓ Chapter Quiz
2 questions

5. Types of Machine Learning Models

○ Classification vs. regression problems
3m 15s

○ How to build a machine learning model in Python
5m 55s

○ Chapter Quiz
2 questions


Conclusion

Machine Learning with Python: Foundations
Chapter Quiz

728

12,568

+



You answered 2 of 2 questions correctly.
You successfully completed all questions in this quiz.

Review all answers

Continue watching

Chapter 5: Types of machine Learning model

ROUGH COPY@2022 - Google D... Documents/TERM2/BDM 1034 x 05_02e - Jupyter Notebook x 04_06e - Jupyter Notebook x +

localhost:8888/notebooks/Documents/TERM2/BDM%201034%20-%20Application%20Design%20for%20Big%20Data/...

jupyter 05_02e (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

The model coefficients correspond to the order in which the independent variables are listed in the training data. This means that the equation for the fitted regression line can be written as:
$$y = 3800.68 + 80.35 \times \text{temperature} - 4665.74 \times \text{humidity} - 196.22 \times \text{windspeed}$$

With the linear regression equation, we can estimate what our model will predict given any weather condition. For example, given a temperature of 72° F, 22% humidity and windspeed of 5 miles per hour, our model would predict:
$$7,578 \text{ bikes} \approx 3800.68 + 80.35 \times 72 - 4665.74 \times .22 - 196.22 \times 5$$

5. Evaluate the Model

In [14]: model.score(x_test, y_test)

Out[14]: 0.9820623857913312

In [15]: y_pred = model.predict(x_test)

In [16]: from sklearn.metrics import mean_absolute_error
mean_absolute_error(y_test, y_pred)

Out[16]: 194.31620720519683

In []:

Chapter Quiz:

Contents

Introduction

1. Machine Learning

2. Collecting Data for Machine Learning

3. Understanding Data for Machine Learning

4. Preparing Data for Machine Learning

5. Types of Machine Learning Models

✓ Classification vs. regression problems
3m 15s

✓ How to build a machine learning model in Python
5m 55s

✓ Chapter Quiz
2 questions


Conclusion

○ Next steps with applied machine learning
1m 25s

Machine Learning with Python: Foundations

Chapter Quiz

728 12,568 + ↗



You answered 2 of 2 questions correctly.

You successfully completed all questions in this quiz.

[Review all answers](#)

[Continue watching](#)

Course Completion Certificate:

