



COLLEGE OF COMPUTER AND INFORMATION SCIENCE

Academic Year 2021-2022

IT Practicum

Final Report

Prepared By:

Marc Kristoffer P. Nullas

Submitted to:

Dean Khristian G. Kikuchi

Practicum Engagement Overview

LinkedIn Learning

To complete the required hours needed, I enrolled various courses via the LinkedIn Learning platform, this was divided into two main modules namely Required Learning and Specialization Learning Module, the hours that was rendered for the required learning module was 135 hours while on the specialization learning module was 189 hours in which the total is 324 hours. The required learning module consists of design thinking, project management, and workplace management module. For the specialization module, it consists of software engineering, machine learning, and data science. I chose topics that can help me understand how data analysis works, whether on data cleaning, data sorting, and data visualization.

Synthesis

I was able to learn about how professional workspaces works, I learned that being a part of a team is great on self development also, because at first i thought that you have to put on a face everytime you go to a workplace in order for you to be a part of them, but here, I learned that speaking up can be beneficial not only you but to your colleagues also. I learned time management here also in which it taught me on how to manage my time well. I do this now by putting a schedule on what I could do for the time being, I now tend to focus on the now not on the tomorrow. Having a schedule is great because you can adjust other things should an emergency come, what I really liked about the first module is that it taught me on how to respect each other's time well. Especially now that we are in the new normal, certain distractions can affect others differently so you have to adapt to it. I have now set an assigned time or date to check emails so that it won't interfere with what I was doing that day. In the second module, I have to learn everything from scratch because it was not taught in my curriculum. This was the python language. I now know basic python and can read them when it is being taught in the movies that I watch to further my knowledge on the data analysis field. I plan to get a google data analytics certificate after graduating to boost my portfolio when applying for a job.

Upon learning the python language, I did the tasks that were asked on the movies and these were on selecting a data set in order to put an analysis on it. There was also machine learning in which I used the titanic data set in order to predict the number of survivors. The speaker said to use IBM SPSS Statistics program wherein you can drag and drop functionalities in order for the machine to predict and learn. I learned that, the more data you can have the more tests the machine can do, and the more tests the machine can do, the more it can predict. There

were many types of algorithms that were explained including naive bayes, k-means and so much more. I also learned the different types of learning these are, supervised, unsupervised and, semi-supervised learning. I was introduced to jupyter notebook as well, this is the main IDE that was used in the movies because it has the libraries it needs such as pandas, numpy, matplotlib, seaborn and so much more. These libraries help with data cleaning, data sorting, and data visualization, and data analysis.

I learned that by using the libraries that were discussed I can now send reports or visualize so that the business analyst can predict what will happen in the future, for example on the movies data set. I used correlation to show that higher budget means higher gross pay. These were extremely hard at first but patience can go a long way. I understand that I need to do more work in order for me to be comfortable in using python on data visualization. There are some tools that were introduced to me such as Power BI, Tableau, and Google Excel. I plan to learn some of these prior to applying for a job since these are what the companies are looking for as a data analyst.

The practicum helped me realize what I was interested in and these were data analysis. It also helped me learn how to read people or read their movements when speaking to them in the workplace.

Movie Portfolio

```
In [1]: import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
plt.style.use('ggplot')
from matplotlib.pyplot import figure
import seaborn as sns

%matplotlib inline
matplotlib.rcParams['figure.figsize'] = (12,8)

In [4]: movies = pd.read_csv('./movies.csv')
```

Figure 1.0 Libraries Used

The dataset that I will be using is from kaggle in which there are a bunch of dataset that can be used in my data analysis journey. This data set will be used for me in order to find the correlation between the budget and the gross earnings. It is interesting for me to learn about how a budget can impact the income that will be generated once the movie goes into cinemas. What I did here first is get the data set from the folder in which I used the pandas library in order for it to help me visualize the data. The figure above shows the attributes and the data corresponding to it. In the next figures I will show how to check if there are any missing values in the data frame.

```
In [7]: movies.isnull().sum()
Out[7]: name      0
rating     77
genre      0
year       0
released    2
score      3
votes       3
director    0
writer      3
star        1
country     3
budget     2171
gross      189
company     17
runtime      4
dtype: int64
```

Figure 1.1 Checking for null values

It shows here that some values are missing, I now have to drop all the values that are missing because it will affect the scatterplot that will be used later.

```
In [9]: newmovies = movies.dropna(axis = 0)
```

```
In [10]: newmovies.isnull().sum()
```

```
Out[10]: name      0  
rating     0  
genre      0  
year       0  
released   0  
score      0  
votes      0  
director   0  
writer     0  
star       0  
country    0  
budget     0  
gross      0  
company    0  
runtime    0  
dtype: int64
```

Figure 1.2 Dropping Null Values

What I've done here next is to make the data frame drop the missing values by using the `isnull()` function by summing them. I am now confident that there will be no values missing. I can now proceed to data cleaning.

In [13]: newmovies

Out[13]:

votes	director	writer	star	country	budget	gross	company	runtime
00.0	Stanley Kubrick	Stephen King	Jack Nicholson	United Kingdom	19000000.0	46998772.0	Warner Bros.	146.0
00.0	Randal Kleiser	Henry De Vere Stacpoole	Brooke Shields	United States	4500000.0	58853106.0	Columbia Pictures	104.0
00.0	Irvin Kershner	Leigh Brackett	Mark Hamill	United States	18000000.0	538375067.0	Lucasfilm	124.0
00.0	Jim Abrahams	Jim Abrahams	Robert Hays	United States	3500000.0	83453539.0	Paramount Pictures	88.0
00.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000.0	39846344.0	Orion Pictures	98.0
...
00.0	Adil El Arbi	Peter Craig	Will Smith	United States	90000000.0	426505244.0	Columbia Pictures	124.0
00.0	Jeff Fowler	Pat Casey	Ben Schwartz	United States	85000000.0	319715683.0	Paramount Pictures	99.0
00.0	Stephen Gaghan	Stephen Gaghan	Robert Downey Jr.	United States	175000000.0	245487753.0	Universal Pictures	101.0
00.0	Chris Sanders	Michael Green	Harrison Ford	Canada	135000000.0	111105497.0	20th Century Studios	100.0
00.0	Hu Guan	Hu Guan	Zhi-zhong Huang	China	80000000.0	461421559.0	Beijing Diqi Yinxiang Entertainment	149.0

Figure 1.3 Raw Data

What I've noticed here now is that both the budget and gross attribute is on the float data type so it results in a .0 result. In order for me to change that i will access the data frame and then change it using astype. On the next figure the results will be shown

```
In [18]: ┌ newmovies['budget'] = newmovies['budget'].astype('int64')
  newmovies['gross'] = newmovies['gross'].astype('int64')
  newmovies['votes'] = newmovies['votes'].astype('int64')
  newmovies['runtime'] = newmovies['runtime'].astype('int64')
|
```

Figure 1.4 Converting attributes to int

By using these codes I have removed the unnecessary values in the attributes namely budget, gross, votes, and runtime. On the next figure I will now show the more presentable data set.

votes	director	writer	star	country	budget	gross	company	runtime
927000	Stanley Kubrick	Stephen King	Jack Nicholson	United Kingdom	19000000	46998772	Warner Bros.	146
65000	Randal Kleiser	Henry De Vere Stacpoole	Brooke Shields	United States	4500000	58853106	Columbia Pictures	104
1200000	Irvin Kershner	Leigh Brackett	Mark Hamill	United States	18000000	538375067	Lucasfilm	124
221000	Jim Abrahams	Jim Abrahams	Robert Hays	United States	3500000	83453539	Paramount Pictures	88
108000	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000	39846344	Orion Pictures	98
...
140000	Adil El Arbi	Peter Craig	Will Smith	United States	90000000	426505244	Columbia Pictures	124
102000	Jeff Fowler	Pat Casey	Ben Schwartz	United States	85000000	319715683	Paramount Pictures	99
53000	Stephen Gaghan	Stephen Gaghan	Robert Downey Jr.	United States	175000000	245487753	Universal Pictures	101
42000	Chris Sanders	Michael Green	Harrison Ford	Canada	135000000	111105497	20th Century Studios	100
3700	Hu Guan	Hu Guan	Zhi-zhong Huang	China	80000000	461421559	Beijing Diqi Yinxiang Entertainment	149

Figure 1.5 Updated Dataframe

Here is the more presentable data set, you can see that the .0 are gone from votes, budget, gross, and runtime.

```
In [38]: plt.scatter(x=newmovies['budget'], y=newmovies['gross'])

plt.title('Budget Vs Gross Earnings')

plt.xlabel('Gross Earnings')

plt.ylabel('Budget for the Film')

plt.show()
```

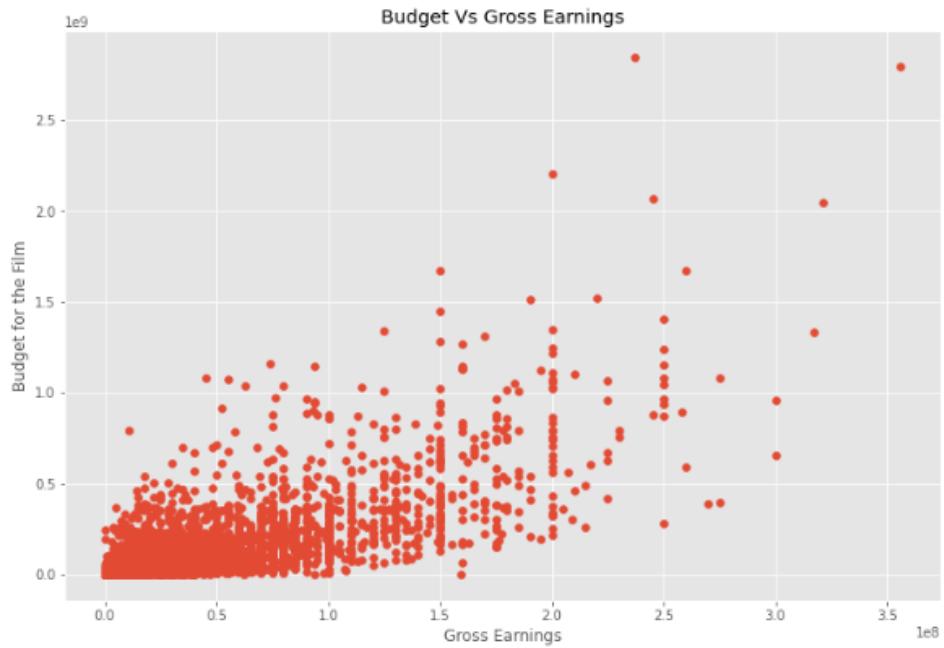


Figure 1.6 Matplotlib Scatterplot

By using matplotlib I was able to visualize some data from the data frame, as you can see there is some correlation happening because it tends to be in an upward trend. In the next figure I will present it with a line using seaborn.

```
In [21]: sns.regplot(x="budget", y="gross", data=newmovies, scatter_kws={"color": "purple"}, line_kws={"color": "black"})
```

```
Out[21]: <AxesSubplot:xlabel='budget', ylabel='gross'>
```

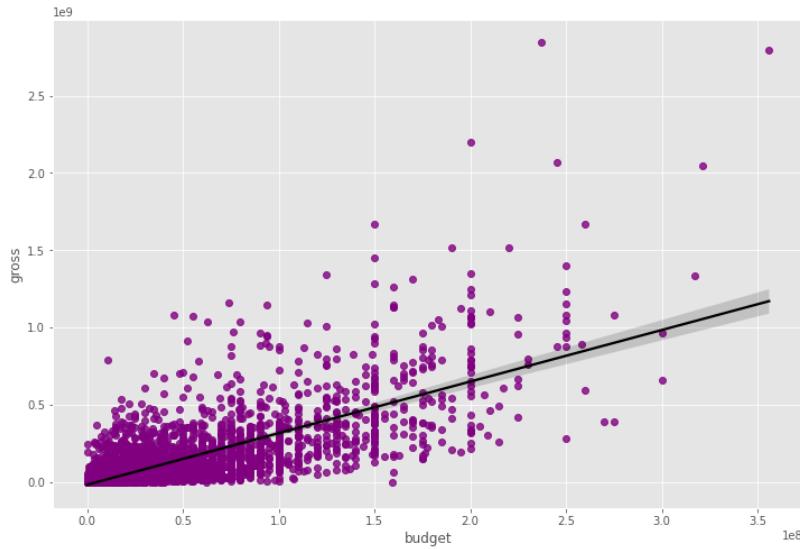


Figure 1.7 Seaborn Scatter Plot

In here I used seaborn to plot it using scatter and then put in a line in order for me to see what trend it is going for. The result is a positive correlation.

```
In [57]: corr_matrix = newmovies.corr(method='pearson')
```

```
corr_matrix
```

```
Out[57]:
```

	year	score	votes	budget	gross	runtime
year	1.000000	0.056386	0.206021	0.327722	0.274321	0.075077
score	0.056386	1.000000	0.474256	0.072001	0.222556	0.414068
votes	0.206021	0.474256	1.000000	0.439675	0.614751	0.352303
budget	0.327722	0.072001	0.439675	1.000000	0.740247	0.318695
gross	0.274321	0.222556	0.614751	0.740247	1.000000	0.275796
runtime	0.075077	0.414068	0.352303	0.318695	0.275796	1.000000

Figure 1.8 Correlation using pearson method

Now this is the correlation between the numeric attributes, in the next figures I will show a more visualized data by using a heatmap matrix on seaborn

```
In [51]: sns.heatmap(corr_matrix, annot=True)

plt.title('Correlation Matrix for Numeric Attributes')

plt.xlabel('Movie Features')

plt.ylabel('Movie Features')

plt.show()
```

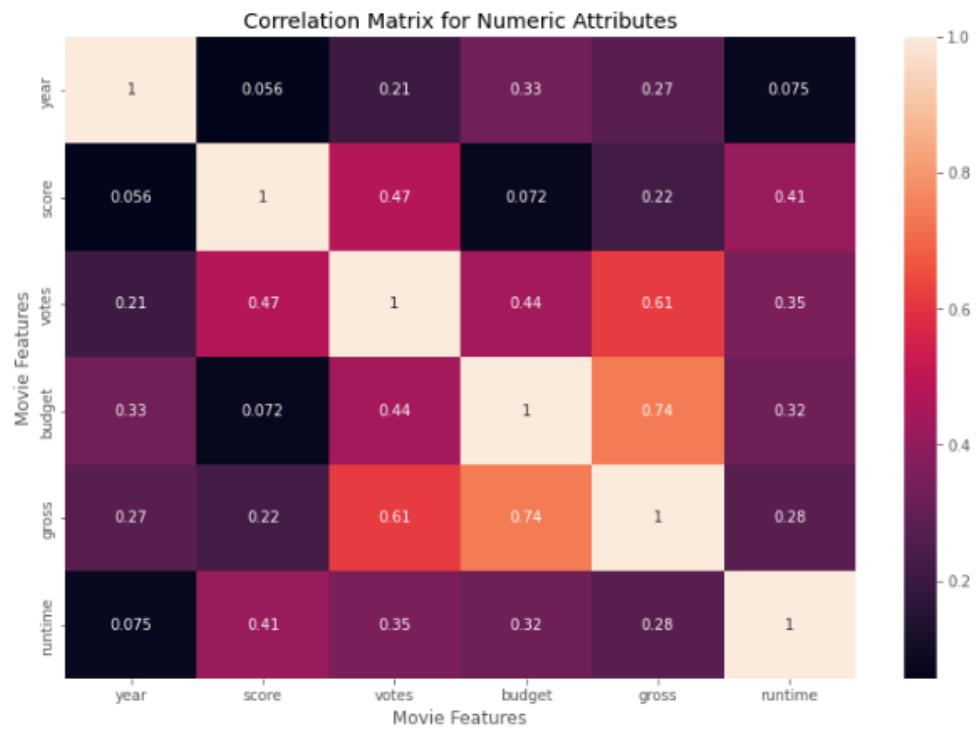


Figure 1.9 Heatmap of Numeric Attributes of the Dataset

Here the data is more visualized and it says here that the brighter the colors are the higher correlation is between the values.

```
In [53]: corr_pairs = corr_matrix.unstack()
sorted_pairs = corr_pairs.sort_values()
sorted_pairs
```

```
Out[53]: year      score      0.056386
          score      year      0.056386
          budget     score      0.072001
          score      budget     0.072001
          year       runtime   0.075077
          runtime    year      0.075077
          year       votes     0.206021
          votes      year      0.206021
          gross      score     0.222556
          score      gross     0.222556
          year       gross     0.274321
          gross      year      0.274321
          gross      runtime   0.275796
          runtime    gross     0.275796
          budget     runtime   0.318695
          runtime    budget    0.318695
          year       budget    0.327722
          budget     year      0.327722
          votes      runtime   0.352303
          runtime    votes     0.352303
          votes      score     0.414068
          score      runtime   0.414068
          votes      budget    0.439675
          budget     votes     0.439675
          votes      score     0.474256
          score      votes     0.474256
          gross      votes     0.614751
          votes      gross     0.614751
          budget     gross     0.740247
          gross      budget    0.740247
          year       year      1.000000
          budget     budget    1.000000
          votes      votes     1.000000
          score      score     1.000000
          gross      gross     1.000000
          runtime   runtime   1.000000
          dtype: float64
```

Figure 2.0 Paired Correlation

Here I've displayed the correlation between them by pairing them, in the next figure I will show only the highest correlation between the numeric values.

```
In [55]: high_corr = sorted_pairs[(sorted_pairs) > 0.5]
```

```
high_corr
```

```
Out[55]: gross      votes      0.614751
          votes      gross      0.614751
          budget     gross      0.740247
          gross      budget     0.740247
          year       year      1.000000
          budget     budget     1.000000
          votes      votes      1.000000
          score      score      1.000000
          gross      gross      1.000000
          runtime   runtime     1.000000
          dtype: float64
```

Figure 2.1 Sorted Pairs

Here I've sorted them all and I can conclude that both votes and budget have a high correlation on gross earnings.

```
cleaned_crime['OFFENSE_CODE_GROUP'].unique()

[: array(['Medical Assistance', 'Towed', 'Offenses Against Child / Family',
       'Simple Assault', 'Investigate Property', 'Verbal Disputes',
       'Larceny', 'Fire Related Reports', 'Missing Person Reported',
       'Robbery', 'Firearm Violations', 'Residential Burglary',
       'Investigate Person', 'Drug Violation', 'Aggravated Assault',
       'Violations', 'Motor Vehicle Accident Response', 'Property Found',
       'Larceny From Motor Vehicle', 'Liquor Violation', 'Other',
       'Harassment', 'Vandalism', 'Property Lost',
       'Assembly or Gathering Violations', 'Warrant Arrests',
       'Landlord/Tenant Disputes', 'Restraining Order Violations',
       'Recovered Stolen Property', 'Auto Theft Recovery',
       'Firearm Discovery', 'Other Burglary', 'Auto Theft',
       'Commercial Burglary', 'Service', 'Police Service Incidents',
       'Disorderly Conduct', 'Property Related Damage',
       'Harbor Related Incidents', 'Fraud', 'Ballistics',
       'Missing Person Located', 'Search Warrants', 'Confidence Games',
       'Counterfeiting', 'License Violation', 'HOME INVASION',
       'Operating Under the Influence', 'License Plate Related Incidents',
       'Evading Fare', 'Embezzlement', 'Prisoner Related Incidents',
       'Aircraft', 'Arson', 'Explosives', 'Homicide', 'Prostitution',
       'Criminal Harassment', 'Bomb Hoax', 'Phone Call Complaints',
       'HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE', 'Manslaughter',
       'HUMAN TRAFFICKING', 'Biological Threat', 'Gambling',
       'INVESTIGATE PERSON', 'Burglary - No Property Taken'], dtype=object)
```

Figure 3.0 Checking for misspellings

Another process of data cleaning is checking for misspellings, I did this by using the unique function as shown on figure 3. Since there are no misspellings the data is now ready to be presented.

Sorting Data

```
In [53]: ca_grouped = ca_babies.sort_values('Count', ascending=False).groupby(['Year', 'Gender']).ca_grouped
```

	Id	Name	State	Count	
Year	Gender				
1910	F	356018	Mary	CA	295
	M	565480	John	CA	237
1911	F	356251	Mary	CA	390
	M	565610	John	CA	214
1912	F	356502	Mary	CA	534
...
2012	M	701517	Jacob	CA	2971
2013	F	557545	Sophia	CA	3451
	M	704421	Jacob	CA	2879
2014	F	561496	Sophia	CA	3162
	M	707312	Noah	CA	2761

210 rows × 4 columns

Figure 4.0

In this task, I have to sort the most popular name and the corresponding year and gender as shown in figure four..

	year	artist	track	time	date.entered	wk1	wk2	wk3	wk4	wk5	...	wk67	wk
0	2000	2 Pac	Baby Don't Cry (Keep...)	4:22	2000-02-26	87	82.0	72.0	77.0	87.0	...	NaN	N
1	2000	2Ge+her	The Hardest Part Of ...	3:15	2000-09-02	91	87.0	92.0	NaN	NaN	...	NaN	N
2	2000	3 Doors Down	Kryptonite	3:53	2000-04-08	81	70.0	68.0	67.0	66.0	...	NaN	N
3	2000	3 Doors Down	Loser	4:24	2000-10-21	76	76.0	72.0	69.0	67.0	...	NaN	N
4	2000	504 Boyz	Wobble Wobble	3:35	2000-04-15	57	34.0	25.0	17.0	17.0	...	NaN	N
...
312	2000	Yankee Grey	Another Nine Minutes	3:10	2000-04-29	86	83.0	77.0	74.0	83.0	...	NaN	N
313	2000	Yearwood, Trisha	Real Live Woman	3:55	2000-04-01	85	83.0	83.0	82.0	81.0	...	NaN	N
314	2000	Ying Yang Twins	Whistle While You Tw...	4:19	2000-03-18	95	94.0	91.0	85.0	84.0	...	NaN	N
315	2000	Zombie Nation	Kernkraft 400	3:30	2000-09-02	99	99.0	NaN	NaN	NaN	...	NaN	N
316	2000	matchbox twenty	Bent	4:12	2000-04-29	60	37.0	29.0	24.0	22.0	...	NaN	N

317 rows × 81 columns

Figure 4.1

Another example is this dataset in which I need to clean up. There is a function in pandas that is called melt, wherein I can combine some variables in order for it to be less heavy when importing it on a database, in the next figure I will demonstrate the function called melt.

```
In [12]: pd.melt(  
    billboard, id_vars=['year','artist','track','time','date.entered'],  
    var_name='weeks',value_name='ratings')|
```

Out[12]:

	year	artist	track	time	date.entered	weeks	ratings
0	2000	2 Pac	Baby Don't Cry (Keep...)	4:22	2000-02-26	wk1	87.0
1	2000	2Ge+her	The Hardest Part Of ...	3:15	2000-09-02	wk1	91.0
2	2000	3 Doors Down	Kryptonite	3:53	2000-04-08	wk1	81.0
3	2000	3 Doors Down	Loser	4:24	2000-10-21	wk1	76.0
4	2000	504 Boyz	Wobble Wobble	3:35	2000-04-15	wk1	57.0
...
24087	2000	Yankee Grey	Another Nine Minutes	3:10	2000-04-29	wk76	NaN
24088	2000	Yearwood, Trisha	Real Live Woman	3:55	2000-04-01	wk76	NaN
24089	2000	Ying Yang Twins	Whistle While You Tw...	4:19	2000-03-18	wk76	NaN
24090	2000	Zombie Nation	Kernkraft 400	3:30	2000-09-02	wk76	NaN
24091	2000	matchbox twenty	Bent	4:12	2000-04-29	wk76	NaN

24092 rows × 7 columns

Figure 4.2 Melted Dataset

In this figure, I have lessened the number of columns from 81 to 7 but made it longer for the rows, this is a good thing because now it has less variables and it makes it easier too for the database programmer to work on. This is still not cleaned because there are still null values, but as you can see it is now more readable and it can help tremendously on predictive modeling.

Visualizing Qualitative Data

```
In [32]: sb.countplot(x='neighbourhood_group', data = listings)
mp.show()
```

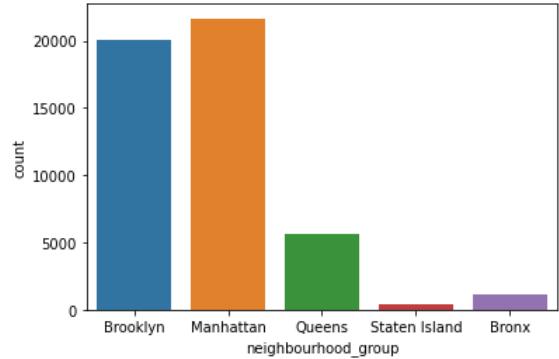


Figure 5.0 Histogram

In this task, I have to show the data by using the libraries matplotlib and seaborn. Matplotlib is a library in python that can help visualize data. Seaborn is also a library that is derived from matplotlib but seaborn is multidimensional. In the figure above, I have to show the count of AirBnB listings from each neighborhood.

HTML Parser

Figure 6.0

In this task, I have to pull contents from the website url using html parser, I did this by using a library called BeautifulSoup. I used the `prettify()` function to better understand the elements of the code block.

```
In [44]: title = title.strip()
        price = price.strip()[1:]

        print(title)
        print(price)

Philips SHP9500 HiFi Precision Stereo Over-Ear Headphones / SHP 9500 Headph
one with Mic Pack
3,699.00
```

Figure 6.1 Stripping Data of Unnecessary Texts

This was the result of me getting the contents of that specific element. I used the strip function to get rid of the Peso sign and the white space of the title of the product.

```
In [55]: today = datetime.date.today()

header = ['Title', 'Price', 'Date']
data = [title, price, today]

with open('LazadaWebScraper.csv', 'w', newline='', encoding='UTF8') as f:
    writer = csv.writer(f)
    writer.writerow(header)
    writer.writerow(data)
```

Figure 6.2 Writing Necessary attributes

Here, I created a csv file and the data that was populated was saved there.

```
In [57]: with open('LazadaWebScraper.csv', 'a+', newline='', encoding='UTF8') as f:
    writer = csv.writer(f)
    writer.writerow(data)
```

Figure 6.3 Appending new data

Next is appending, by doing so, the csv will now be populated each time I run the code block.

```
In [70]: df = pd.read_csv(r'./LazadaWebScraper.csv')

print(df)
```

Figure 6.4 Reading new data

By doing this I can now see if the product I want changes its price by the timer ends, the limitation of this is that it is not running when the machine is turned off since it is not yet in the cloud.

			PRICE	TITLE	DATE
0	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
1	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
2	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
3	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
4	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
5	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
6	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
7	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
8	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
9	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
10	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
11	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
12	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
13	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
14	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
15	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
16	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
17	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
18	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
19	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
20	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
21	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
22	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
23	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
24	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
25	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		
26	Philips SHP9500 HiFi Precision Stereo Over-Ear...	3,699.00	2022-07-12		

Figure 6.4 Populated dataset

This is the sample result of it populating. This is only one product for now but all sections of the website can be scraped using the id elements of the web page.

Logistic Regression

```
In [6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
--- 
 0   age        303 non-null   int64  
 1   sex        303 non-null   int64  
 2   cp         303 non-null   int64  
 3   trestbps  303 non-null   int64  
 4   chol       303 non-null   int64  
 5   fbs        303 non-null   int64  
 6   restecg   303 non-null   int64  
 7   thalach   303 non-null   int64  
 8   exang     303 non-null   int64  
 9   oldpeak   303 non-null   float64 
 10  slope      303 non-null   int64  
 11  ca         303 non-null   int64  
 12  thal       303 non-null   int64  
 13  target     303 non-null   int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

Figure 7.0

In order for me to perform logistic regression, the data set must be binary. It means that it will be answerable by yes and no. This data set came from kaggle and it contains the target attribute that says that if the result is 1 the heart is not healthy while having 0 means the heart is healthy. This data set will now be used on prediction using LogisticRegression.

```
In [10]: df['target'].value_counts()
```

```
Out[10]: 1    165  
0    138  
Name: target, dtype: int64
```

```
In [13]: x = df.drop(columns='target', axis=1)  
y = df['target']
```

```
In [14]: x
```

```
Out[14]:
```

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2
...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2

303 rows × 13 columns

Figure 7.1 Splitting training and test data set

I have now separated the column target from the data set which means I can now split the data into training and test sets by using sklearn train_test_split().

```
In [16]: x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.2, stratify=y, random_state=2)
```

```
In [17]: print(x.shape, x_train.shape, x_test.shape)  
(303, 13) (242, 13) (61, 13)
```

Figure 7.2 Shaping the model

I now began to separate the test and the training values. According to the videos about machine learning you can split the data into two resulting in test and training whether it be 10% test data and 90% training data or 20% test data or 80% training data. In this case I've used the 80/20 principle. Now I can train the model using logistic regression

```
In [18]: model = LogisticRegression()

In [19]: model.fit(x_train, y_train)

C:\Users\MK\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
63: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

    Increase the number of iterations (max_iter) or scale the data as shown in:
        https://scikit-learn.org/stable/modules/preprocessing.html
    Please also refer to the documentation for alternative solver options:
        https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
        n_iter_i = _check_optimize_result()

Out[19]: LogisticRegression()

In [20]: x_predict = model.predict(x_train)
accuracy_training_data = accuracy_score(x_predict, y_train)

In [21]: print('Accuracy on Training Data: ', accuracy_training_data)

Accuracy on Training Data:  0.8512396694214877

In [22]: x_predict_test = model.predict(x_test)
accuracy_test_data = accuracy_score(x_predict_test, y_test)

In [23]: print('Accuracy on Test Data: ', accuracy_test_data)

Accuracy on Test Data:  0.819672131147541
```

Figure 7.3 Logistic Regression Model

Here I've fitted the model, this will be the percentage of accuracy of the regression on the test and training data. On the training data I got 85% and on the test data I got almost 82%.

```
In [25]: # input_data = (62,0,0,140,268,0,0,160,0,3.6,0,2,2)

input_data_as_numpy_array= np.asarray(input_data)

input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if (prediction[0]== 0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')

[0]
The Person does not have a Heart Disease
```

Figure 7.4 Predictive System

Here I copied one row from the csv file to see if the model is predicting correctly. In this case the row resulted in 0 which means the person is healthy and the predictive model also predicted that the person has a healthy heart.

Competency-based Resume

Marc Kristoffer Nullas

kristoffnullas@gmail.com
(+63)9694617275
San Pedro, Laguna

Summary

I am results-oriented, constantly checking in with the goal to determine how close or how far away we are and what it will take to make it happen. I find this pressure inspiring and a great motivator for the rest of the team.

Online links

<https://www.linkedin.com/in/marc-kristoffer-nullas-3b7305189/>
<https://github.com/mkpnallas>

Education

Bachelor of Science in Computer Science

Malayan Colleges Laguna • Cabuyao, Laguna

Languages

Python, HTML, CSS, JavaScript

Databases

SQL, MySQL

Software

Jupyter Notebook, Google Excel, Visual Studio, PhpMyAdmin

Projects

- Movie Correlation of budget and gross earnings
 - Lazada Web Scraper
 - Designing A Web Based DGBL with Intelligent Tutoring System for Vocabulary Enhancement of College Students (Data Visualization of Learning Gain)
 - Heart Disease Prediction using Logistic Regression
-

Certificates

LinkedIn

- Python for Data Science Essential Training
- Python Data Analysis
- Python Essential Training
- Programming Foundations: Software Testing/QA
- Machine Learning with Python: Foundations
- Learning the Python 3 Standard Library

Practicum Learning Path Module Proposal



Malayan Colleges Laguna
A MAPÚA SCHOOL



College of Computer
and Information Science
Malayan Colleges Laguna

IT PRACTICUM X LEARNING PATH PROPOSAL

COMPLETE NAME : Nullas, Marc Kristoffer P.
STUDENT NUMBER : 2013100405 PROGRAM : CS

IDENTIFIED LEARNING MODULES

A. Required Learning Modules

1. Design Thinking Learning Module (*Total Engagement Hours: 15 hours Total Credit Hours: 45 hours*)
 - a. Learning Design Thinking: Lead Change in Your Organization by Turi McKinley (2h 8m)
 - b. Develop Your Creative Thinking and Innovation Skills Learning Path 8 courses (8h)
 - c. Develop Critical-Thinking, Decision-Making, and Problem-Solving Skills 6 courses (5h)
2. Project Management Learning Module (*Total Engagement Hours: 15 hours Total Credit Hours: 45 hours*)
 - a. Project Management Foundations by Bonnie Biafore (3h 20m)
 - b. Become an Agile Project Manager Learning Path 9 courses (12h)
3. Workplace Management Learning Module (*Total Engagement Hours: 15 hours Total Credit Hours: 45 hours*)
 - a. Develop Your Communication Skills and Interpersonal Influence 9 courses (8h)
 - b. 5S Workplace Productivity by Gemba Academy (1h 36m)
 - c. Time Management Fundamentals by Dave Crenshaw (2h 53m)
 - d. Time Management: Working from Home by Dave Crenshaw (1h 25m)
 - e. Business Etiquette: Phone, Email, and Text by Suzanna Kaye (58m)



B. Specialization Learning Modules

1. Software Engineering (*Total Engagement Hours: 27 hours Total Credit Hours: 81 hours*)

- a. Software Design: From Requirements to Release by Neelam Diwedi (1h 54m)
- b. Programming Foundations: Software Testing/QA by Michele Vallisneri (53m)
- c. Agile Software Development by Shashi Shekhar (1h 53m)
- d. Python for Engineers and Scientists by Michele Vallisneri (1h 53m)
- e. DevOps Foundations by Ernest Mueller and James Wickett (2h47m)
- f. Succeeding in DevOps Laura Stone (32m 12s)
- g. AWS Essential Training for Developers by Jeremy Villeneuve (3h 17m)
- h. Python Essential Training by Bill Weinman (4h 37m)
- i. Python Quick Start by Lavanya Vijayan, Madecraft (1h 36m)
- j. Python for Non-Programmers by Nick Walter (1h 55m)
- k. Learning Python by Joe Marini (3h 10m)
- l. Learning the Python 3 Standard Library by Kathryn Hodge (2h 9m)

2. Machine Learning (*Total Engagement Hours: 21 hours Total Credit Hours 63 hours*)

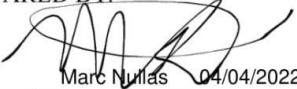
- a. Artificial Intelligence Foundations: Machine Learning by Doug Rose (1h 17m)
- b. AWS Machine Learning by Example by Jonathan Fernandes (1h 25m)
- c. Machine Learning with Python: Foundations by Frederick Nwanganga (1h 43m)
- d. Become a Machine Learning Specialist by Keith McCormick, Dan Sullivan, Lillian Pierson, P.E., Adam Geitgey (17h 18m)



3. Data Science (*Total Engagement Hours: 15 hours Total Credit Hours: 45 hours*)

- a. Introduction to Data Science by Lavanya Vijayan (1h 48m)
- b. Learning Data Science: Understanding the Basics by Doug Rose (1h 16m)
- c. Python Data Analysis by Michele Vallisneri (2h 30m)
- d. Python for Data Science Essential Training Part 1 by Lillian Pierson, P.E. (6h 2m)
- e. Python for Data Science Essential Training Part 2 by Lillian Pierson, P.E. (3h 44m)

PREPARED BY:



Marc Nullas 04/04/2022

Signature over Printed Name / Date

APPROVED BY:



Khristian G. Kikuchi

Signature over Printed Name / Date

Mapúa Institute of Technology at Laguna · College of Computer and Information Science · E.T. Yuchengco College of Business · Mapúa-PTC College of Maritime Education and Training
College of Arts and Science · Institute for Excellence in Continuing Education and Lifelong Learning · MCL Senior High School

Address : Cabuyao, Laguna 4025
Trunkline : +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 520-8975
Email : mclinfo@mcl.edu.ph

www.mcl.edu.ph  MalayanColleges   MalayanLaguna

Weekly Journals



Malayan Colleges Laguna
A MAPUA SCHOOL

REVISION NO.: 00
REVISION DATE: May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
- SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
- HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	4/22/2022	AREA ASSIGNMENT	NA
TASK	Software Engineering/ Workplace Management Learning Module	SHIFT/TIME	NA

The courses i've learned here are some courses from software engineering and more on the Workplace Management Module, On the software engineering task, i've started learning the basics of python. In here i've learned about the python language in which I will use in the more advanced courses on machine learning and data science. I was not familiar with python language because It was not taught in the curriculum so i had to start from scratch. I did have an easier time understanding it because It follows the same logic but only has a different syntax. It helped me recall on programming also because i was not coding or programming for the past two years due to the pandemic and i had to help in our family business. I've also learned about on how to be a quality assurance manager, here it taught me on how to collaborate with my team more. I understood more that making mistakes is a part of learning and i shouldn't be disappointed by it too much. I've also learnt that by having a stress free environment the team can be more open with themselves and can help with each other more. On the Workplace Management module, i've started learning about time management. In here, It is all about productivity and on how to maximize your time efficiently. The biggest struggle i have right now is the stress of doing all things at once, since I have other commitments outside. I'm pressed on things that are needed to be done for the day so i tend to multitask. In the past, i thought if i just multitask them i will get it done but what i didn't realize is when i multitask it takes a toll on me as well, and it amounts to more stress whereas if I planned it more I will get the same results but the stress will be reduced and it will lead to me maximizing my productivity. It also taught me on how to clear my mind by having a schedule of it by having a break so i can focus more later. This course made me better at managing my time. This course made me better at managing my time and made me understand myself more. One course has an exam in which i got at least 70%.

TRAINEE'S SIGNATURE



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	4/29/2022	AREA ASSIGNMENT	NA
TASK	Workplace Management Learning Module	SHIFT/TIME	NA

Since I learned about time management in the previous course, I implemented some of it so I accomplished more.

Here I learned about time management at home, business etiquettes, workplace productivity, Interpersonal and nonverbal communication.

The course time management at home says that by having a white noise generator, It can help improve my productivity since it isolates the unnecessary noises in my environment

The course has an exam in which I got 89%. Next is the course business etiquettes by Phone, email, and Text. This course has taught me on how to be more professional in sending messages, and on how to professionally follow-up an email by not making it rude.

Next is the 5s workplace productivity, The 5s productivity is a methodology that helps or makes others to be more efficient in the video there was a scenario wherein the speaker interviewed a person in the workplace that is using the methodology. He said that how can i set the next person that uses the terminal to be more successful in his job. This was really interesting to me because in our culture this is quite unusual because we tend to not care about the next person that is going to use this because it is not us. Next is the nonverbal communication, This is quite an interesting find for me because now I can evaluate what the person is feeling or thinking at the moment, I can now see their cues if they are not comfortable. Lastly the interpersonal course, this course guided me on how to work or respond well under pressure. I am somewhat of an introvert and this really helped me on how to manage it.

This course has an exam in which i got 89%



TRAINEE'S SIGNATURE

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
- SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
- HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	5/6/22	AREA ASSIGNMENT	NA
TASK	Design Thinking Learning Module	SHIFT/TIME	NA

At first, I thought being creative means you have to be artistic.

But now that i've finished these modules, I can now say that I am somewhat creative. Here, It says that being creative means you know how to solve problems whether it will be quick or rational. I also now see the world around me differently than what i've used to. It also taught me on how to facilitate group creativity. We do this by asking the team on their inputs and discussing it with the team. I will make sure that the environment is a free thinking space so that all members will be comfortable on dishing out ideas that can help us in the long run. I also saw the benefits on having a team building or a bootcamp wherein the team plays a game that can boost their confidence on each other and by knowing more of the person they are working with. I observed that before they did their bootcamp they are a bit shy to talk to each other because they came from different cultures. After the bootcamp, now the team is more efficient on collaborating with each other to finish the tasks. In the course, Develop Your Creative Thinking and Innovation Skills there was an exam in which I got 80%. It also taught me on how to be an innovator the speaker said that some of the successful businesses started on a crazy idea, and that crazy idea turned into something big. What they did was just ask a question, and that question lead to different answers, and different answers means that now you can establish a foundation on what the customers want or expect on your product. I also found out that, I am an Experimenting type of person. This means that trial and error is the way for me to truly understand something.

In the course, Critical thinking, there was an exam and I got 83%



TRAINEE'S SIGNATURE



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	5/13/22	AREA ASSIGNMENT	NA
TASK	Project Management Learning Module	SHIFT/TIME	NA

This module taught me all i need to know on agile methodology. Since i was used to waterfall i learned that using agile can resolve more issues faster. It taught me all about on how to be a scrum master and on how to facilitate them. The tool that has helped me the most that is taught on this module is Trello. Trello is all about making lists or cards in this matter. Trello helps me on managing my schedules, my to-do lists. And checking my progress on the things that ive been working on. Trello helps on agile methodology by having a backlog and it can set high or low priorities. These helps on prioritizing what needs to be done first. I also learned the roles of a scrum master and a product owner. At first, I thought that by being a scrum master you have to manage everything including the team, but i was wrong. I learned that by being a scrum master you only have to manage the agile framework, eliminate distractions and do reports, checking estimates and is also responsible for encouraging the team. Scrum master also needs to make sure that the team is doing what their respective role is and is also responsible to reward or compliment them individually. Its role is to also make the product owner sit by the development team or make communications easier between them. I thought at first that by being a product owner means you just own and wait for the team to finish them and do some feedback and the product manager sends a report to the development team, These takes more time, but, by using agile, the product owner is directly communicating or helping the development team by making a checklist or a list of priorities to be done first. In doing so, the team now knows what to be done first and what to put on last. Some of the courses has an exam in which i got 80%, 77%, 71%, 74%, 90%, 86%, and 83%

TRAINEE'S SIGNATURE

COPY: (1) STUDENT; (2) PRACTICUM ADVISER

FORM OVPAA 030G

THIS FORM IS AVAILABLE AT THE OVPA.



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	5/20/22	AREA ASSIGNMENT	NA
TASK	Software Engineering/Machine Learning	SHIFT/TIME	NA

I have started some courses in the two modules but i am not finished with them because i needed more time to actually learn them, I am now doing the exercises that are posted in each courses and aim for myself to be more comfortable in using them. In order for me to proceed to the other courses, I've finished some modules in software engineering that will serve as my foundation in order for me to understand what the speaker is saying or doing. What i've learned so far is on data visualization in machine learning by using python and by defining or classifying what the data is, whether it be large or small training sets, or whether it will be supervised, unsupervised, and semi-supervised learning. I also am learning on how to use jupyter notebook, jupyter notebook is a popular or more commonly used environment among machine learning or data scientists. Jupyter notebook has a package called pandas in which it can visualize the data by putting it on a table that has rows and columns, columns being the variables that can be called on the instructions that will be put on each cell block in the notebook. For example, i have a sample data with students details. At first, the data is unfiltered or not clean since some of the values has no data in it. Another thing is some values can be filled by using its median to replace the missing ones. This is really challenging for me since i am not used in the environment. In order for me to resolve this, i aim to study or use them more in order for me to be comfortable by it. My progress currently in software engineering is 50% while on machine learning im only on 10%



TRAINEE'S SIGNATURE



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	5/27/2022	AREA ASSIGNMENT	NA
TASK	Machine Learning	SHIFT/TIME	NA

When i was using jupyter notebook online i noticed that there was no intellisense automatically it has to be manually by pressing tab key, but i researched and found out that I can make it automated by installing the Hinterland extension. I've learned on how to select the proper variables in doing charts meaning I now have the gist on what the independent variable and the dependent variable is. I've learned also about the two statistical model namely CHAID and C&RT. The difference between the two is on how to deal with missing data, CHAID deals with it by grouping them with the one that has a mutual connection, while since C&RT only deals with binary, meaning it only has 1 or 0, or only boolean. It handles it by creating surrogates so that no data will be missing. I've also learned on how to visualize a small data using linear regression, in there, the speaker encouraged to use the IBM SPSS statistics software in which it can help visualize it. I did scatter plots in here in which I studied whether there really is a wage gap between genders, and i was surprised to learn that there is due to a specific reason. Gender is a nominal variable in which being 0 being the male, and 1 being the female. Due to computations, even if both have the same education level, it will never be equal, there will always be a gap. This is because, in code, with the male being equal to 0, the sex now will be 0 meaning it will not be subtracted, while the female being 1, it will be subtracted.

[Signature]

TRAINEE'S SIGNATURE



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	6/3/2022	AREA ASSIGNMENT	NA
TASK	Data Science/Machine Learning	SHIFT/TIME	NA

Since I vaguely know the terms and definitions that are being taught in the machine learning module, I decided to proceed to data science first in order for me to fully understand the data that are being used in the examples, the examples that are used are already cleaned and visualized so I dont know how that worked so i decided to learn data science first before machine learning. When i started the first course in my module It now became clear on what the speaker is talking about while giving examples, i felt like i was back on track so i was really happy that i decided to switch to the data science module first before machine learning, in here I learned more on the pandas library and on how to use them such as the loc method and on how to sort the values in the dataset. I saw this on the machine learning module too, but the speaker didn't elaborate on it too much so i felt lost and confused.

TRAINEE'S SIGNATURE

COPY: (1) STUDENT; (2) PRACTICUM ADVISER

FORM OVPAA 030G

THIS FORM IS AVAILABLE AT THE OVPA.



REVISION NO.:	00
REVISION DATE:	May 10, 2016

DAILY JOURNAL

IMPORTANT INFORMATION

- INCLUDE TASK ASSIGNMENTS OR MOVEMENTS, REFLECTION ON THE DAY'S NEW LEARNING, ACCOMPLISHMENT, CHALLENGES FACED AND HOW YOU RESPONDED, OBSERVATIONS AND RECOMMENDATIONS ON THE IMPROVEMENT OF SYSTEMS / OPERATION / MANAGEMENT, ETC.
 - SCANNED COPIES OF THIS FORM SHALL BE SUBMITTED ON A WEEKLY BASIS THROUGH APPROVED LMS.
 - HARD COPIES OF THIS FORM SHOULD BE COMPILED AS PART OF THE STUDENT'S PORTFOLIO.

DATE	6/10/2022	AREA ASSIGNMENT	NA
TASK	Data Science	SHIFT/TIME	NA

Now that I've finished the last movies, I managed to catch up on what the speaker was saying, I now know how to clean data, check for misspellings, sort them and remove unnecessary columns. Now that I've done that I can now begin to visualize the questions that will come. In the videos the speaker said to use matplotlib, numpy, and seaborn to help me visualize it more. Matplotlib library helps on histograms very well because it has features such as the labeling. On my projects, I used Matplotlib to create a histogram. Numpy helps on arranging the data that will be put on the data frame. Seaborn is useful on heatmaps and scatterplots. Now that I've finished my practicum, I conclude that I really am interested in data analysis and I want to practice more so that I can present the data on the people that will handle such data. Data analysis can help businesses on what to improve on and it can help influence them on the following years that will come.

TRAINEE'S SIGNATURE

COPY: (1) STUDENT; (2) PRACTICUM ADVISER

FORM OVPAA 030G

THIS FORM IS AVAILABLE AT THE OVPA.

Certifications



LinkedIn LEARNING
Certificate of Completion
Congratulations, Marc Kristoffer Nullas

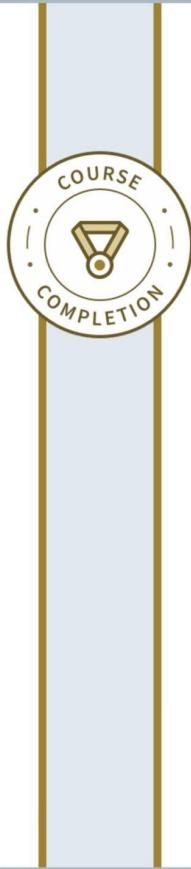
Time Management: Working from Home
Course completed on Apr 26, 2022 at 09:35AM UTC • 1 hour 17 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.


Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AQ4wlcWj_dNRtSiZurlQuJzxLQSN



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Time Management Fundamentals

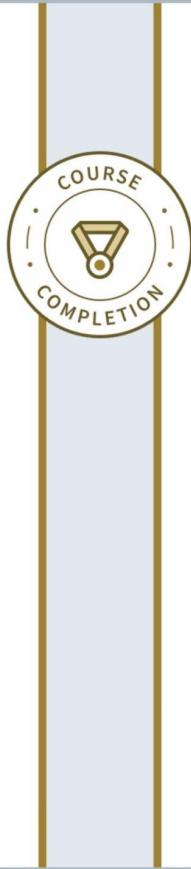
Course completed on Apr 25, 2022 at 08:04AM UTC • 1 hour 45 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AcBEECz4SBLNjGZ27iUvXXGkQp8I



LinkedIn LEARNING

Certificate of Completion
Congratulations, Marc Kristoffer Nullas

Software Design: From Requirements to Release

Course completed on May 11, 2022 at 09:31AM UTC • 1 hour 54 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.



Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Ae7Sjv2vX-lf-rWR8h6o_xM3TNgA



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python Quick Start

Course completed on Apr 20, 2022 at 06:19AM UTC • 1 hour 36 min

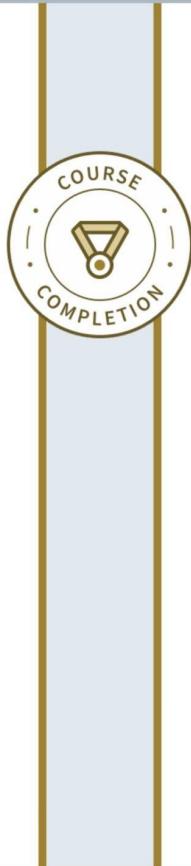
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AWGBNSp8ra0q6ByiPPAk8moWDIN



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python for Non-Programmers

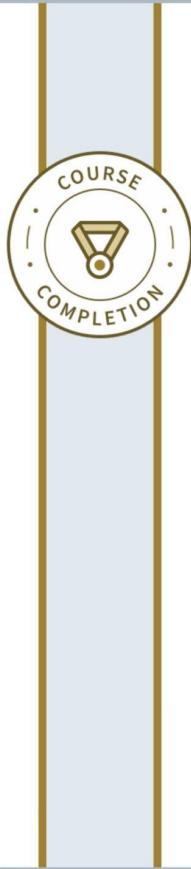
Course completed on Apr 19, 2022 at 02:03PM UTC • 1 hour 55 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AXefrlgtXyY2-a45KRZC6L6ofVts



LinkedIn LEARNING

Certificate of Completion
Congratulations, Marc Kristoffer Nullas

Python for Engineers and Scientists

Course completed on May 03, 2022 at 06:54AM UTC • 1 hour 58 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: ATsDcv2e7y6LBisbqYwTZUqwMJ8v



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python for Data Science Essential Training Part 1

Course completed on Jun 10, 2022 at 07:07AM UTC • 6 hours 2 min

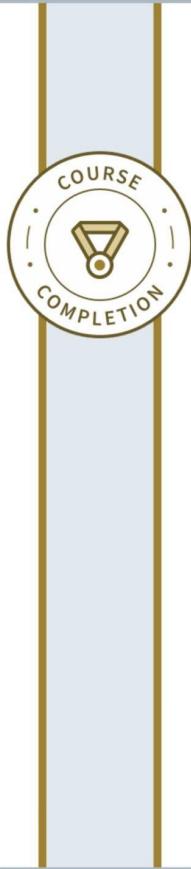
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AWzOqjTT79GmorqOWMVXsuVZB8nE



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python for Data Science Essential Training Part 2

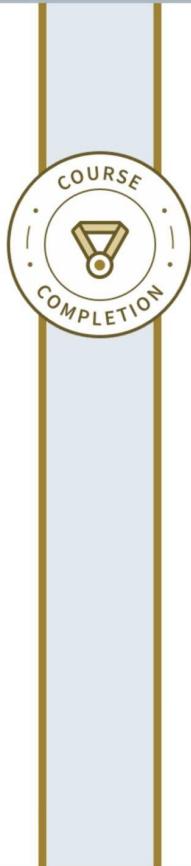
Course completed on Jun 10, 2022 at 11:45AM UTC • 3 hours 44 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AQLRWmiW-YJf9lqUYrJOzR6DrVHF



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python Essential Training (2018)

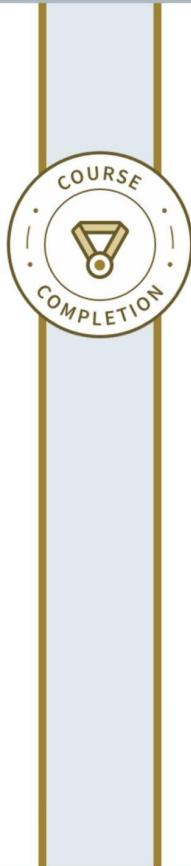
Course completed on Jun 12, 2022 at 07:58AM UTC • 4 hours 37 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AZISNGHlufivBEVDzXUybVTsqZNt



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Python Data Analysis

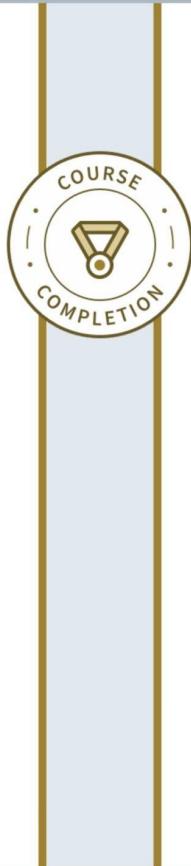
Course completed on Jun 06, 2022 at 01:19PM UTC • 2 hours 30 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AZ_TLw53CQE3_ei3Hv4frlcujDg



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Project Management Foundations

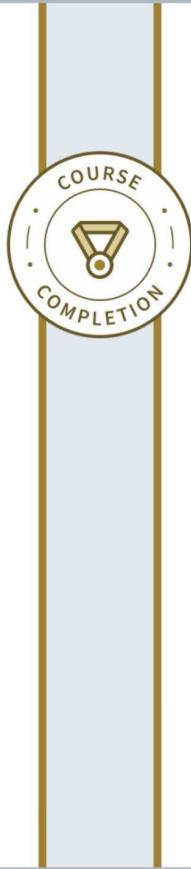
Course completed on May 06, 2022 at 10:25AM UTC • 3 hours 20 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AVXHCzx5KqFetgrEM5Op8qFugf2Y



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Programming Foundations: Software Testing/QA

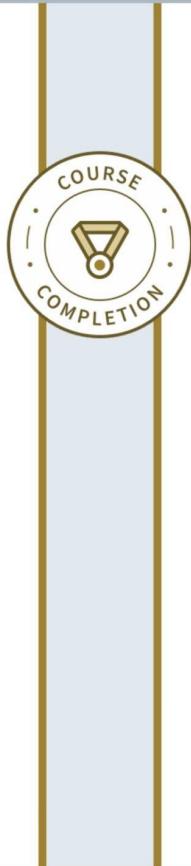
Course completed on Apr 24, 2022 at 05:32AM UTC • 53 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AaDsdsQh79dmfNNn5WXiV8xowFei



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Machine Learning with Python: Foundations

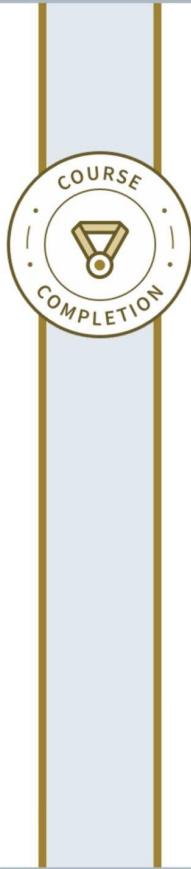
Course completed on May 23, 2022 at 06:06AM UTC • 1 hour 43 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AS_5Q3NKFNdK7KVHMw_gY3jFGPFa



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Learning the Python 3 Standard Library

Course completed on May 11, 2022 at 07:43AM UTC • 2 hours 9 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AV5NJ1Ds0UR0Jd4E1vdFH_kZCEI4



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Learning Python

Course completed on Apr 28, 2022 at 04:09PM UTC • 3 hours 10 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Afrx46_JwwQKeSkfb9eQ7G8ZOvr



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Learning Design Thinking: Lead Change in Your Organization

Course completed on May 02, 2022 at 04:25AM UTC

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink.

Head of Content Strategy, Learning

LinkedIn Learning

1000 W Maude Ave

Sunnyvale, CA 94085

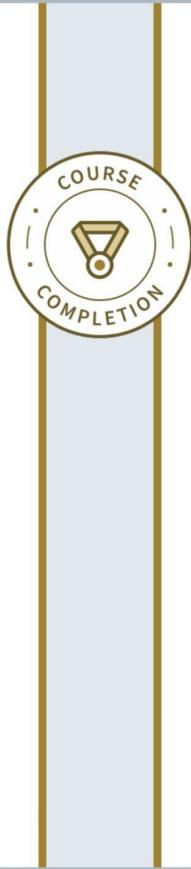


The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101

Certificate No: AROX0rEGfrkp9Fg4pBrlwImMn2w

PDUs/ContactHours: 2.00 | Activity #: 4101CK30K2



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Learning Data Science: Understanding the Basics

Course completed on Jun 06, 2022 at 07:41AM UTC • 1 hour 16 min

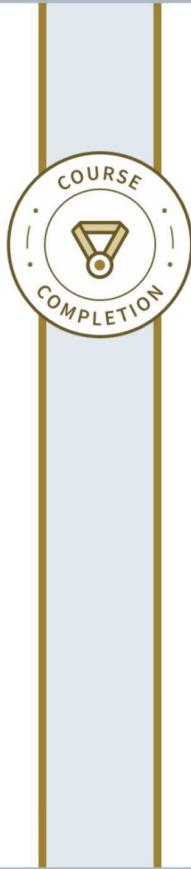
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.



Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Ab6e5fmr4uk4qWc9g0AOoEd92Z2-



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Introduction to Data Science

Course completed on Jun 06, 2022 at 07:32AM UTC • 1 hour 48 min

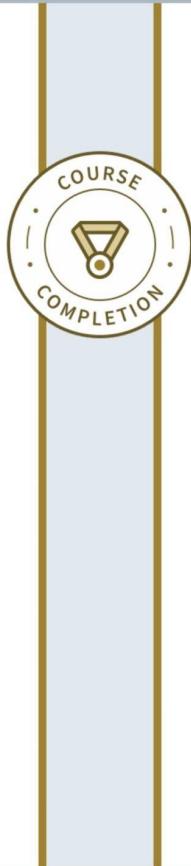
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.



Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AXyutVwmR27M299dmorCzemH6G5a



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

DevOps Foundations

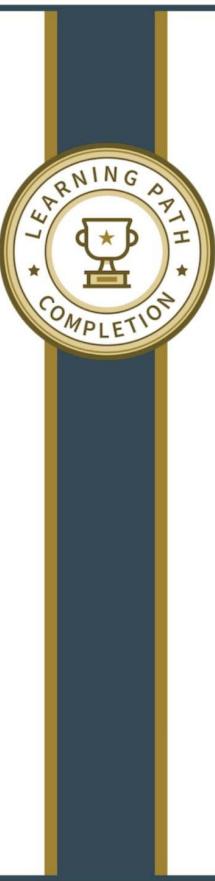
Course completed on May 18, 2022 at 08:16AM UTC • 2 hours 47 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AfyrG6xMqursRF9DarfCbVmniP_



in LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Develop Your Creative Thinking and Innovation Skills

Learning Path completed on May 02, 2022 at 05:08PM UTC • 8 hours 25 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Top skills covered

Organizational Leadership, Creativity Skills, Innovation Management, Collaborative Innovation

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: ASeUav8De94aqX36UQ_1BvLxDhXO



in LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Develop Your Communication Skills and Interpersonal Influence

Learning Path completed on Jul 06, 2022 at 06:48AM UTC • 7 hours 43 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Top skills covered

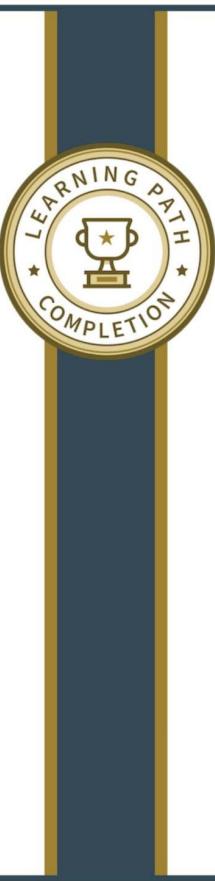
Communication, Interpersonal Skills, Interpersonal Communication

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AZ9FdEjLyppeMEX7poXZh9nFlttD



in LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Develop Critical-Thinking, Decision-Making, and Problem-Solving Skills

Learning Path completed on May 04, 2022 at 11:47AM UTC • 4 hours 52 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Top skills covered

Problem Solving, Critical Thinking, Decision-Making

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AeSZFlcFapyowhfYZTlv25QBCZe



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Business Etiquette: Phone, Email, and Text

Course completed on Apr 26, 2022 at 10:54AM UTC • 58 min

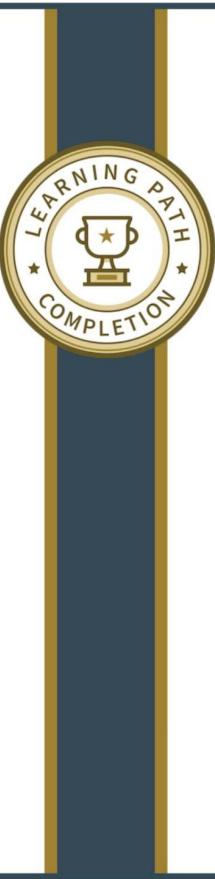
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AVJrYtLpfe3H41TqVuPDA2qMiYTR



in LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Become an Agile Project Manager

Learning Path completed on May 08, 2022 at 08:07AM UTC • 10 hours 34 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Top skills covered

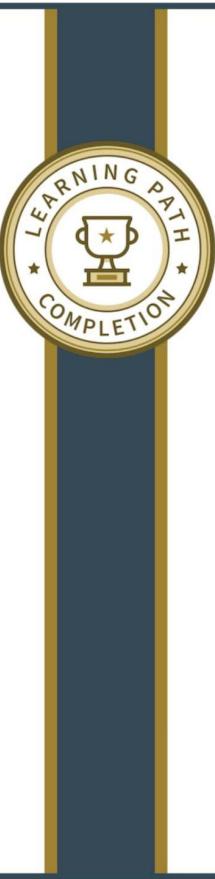
Project Management, Microsoft Project, Agile Project Management

A handwritten signature in black ink.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Ad5fsI18U519CAMLV97GR4z5dEV8



in LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Become a Machine Learning Specialist

Learning Path completed on Jun 12, 2022 at 04:02AM UTC • 17 hours 18 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Top skills covered

Data Science

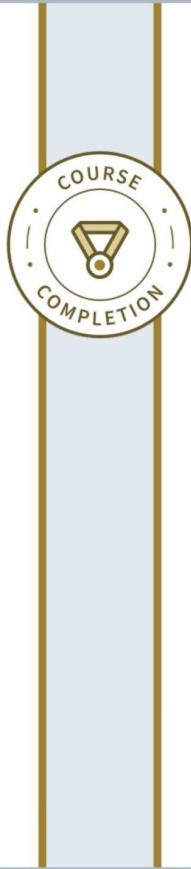
Head of Content Strategy, Learning

LinkedIn Learning

1000 W Maude Ave

Sunnyvale, CA 94085

Certificate Id: AU4dwmovvyI3P8UeDfKwevipGMKm



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

AWS Machine Learning by Example

Course completed on Jun 20, 2022 at 07:54AM UTC • 1 hour 25 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AXM8cCuzrv6aalyCNEdN8xCBXN-L



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

AWS Machine Learning by Example

Course completed on Jun 20, 2022 at 07:54AM UTC • 1 hour 25 min

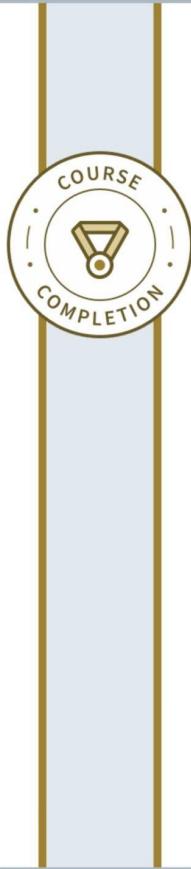
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AXM8cCuzrv6aalyCNEdN8xCBXN-L



LinkedIn LEARNING

Certificate of Completion
Congratulations, Marc Kristoffer Nullas

AWS Essential Training for Developers

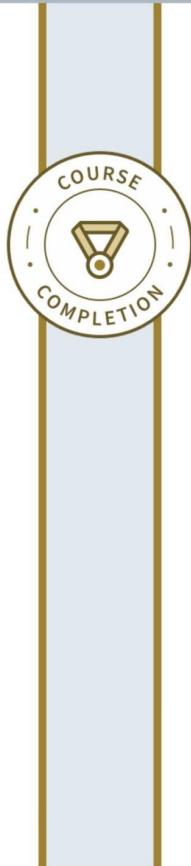
Course completed on May 12, 2022 at 07:14AM UTC • 3 hours 17 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Ad_hhg5CSt8E9Ssgk7U7LhytS-TO



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Artificial Intelligence Foundations: Machine Learning

Course completed on May 19, 2022 at 07:27AM UTC • 1 hour 17 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AfTJbh1bZJuelvJ-Ug8G6D7B5pF8



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

Agile Software Development

Course completed on May 08, 2022 at 10:43AM UTC • 1 hour 53 min

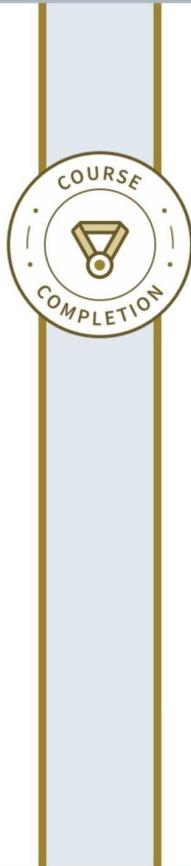
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

A handwritten signature in black ink that reads "Dan Boddy".

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: Abe1wSLIm0By8lieNfx-1xEKjfjF



LinkedIn LEARNING

Certificate of Completion

Congratulations, Marc Kristoffer Nullas

5S Workplace Productivity

Course completed on Apr 26, 2022 at 01:38PM UTC • 1 hour 36 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Head of Content Strategy, Learning

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AYXkEOKQMFmoqFf4CHRxsazAnNL