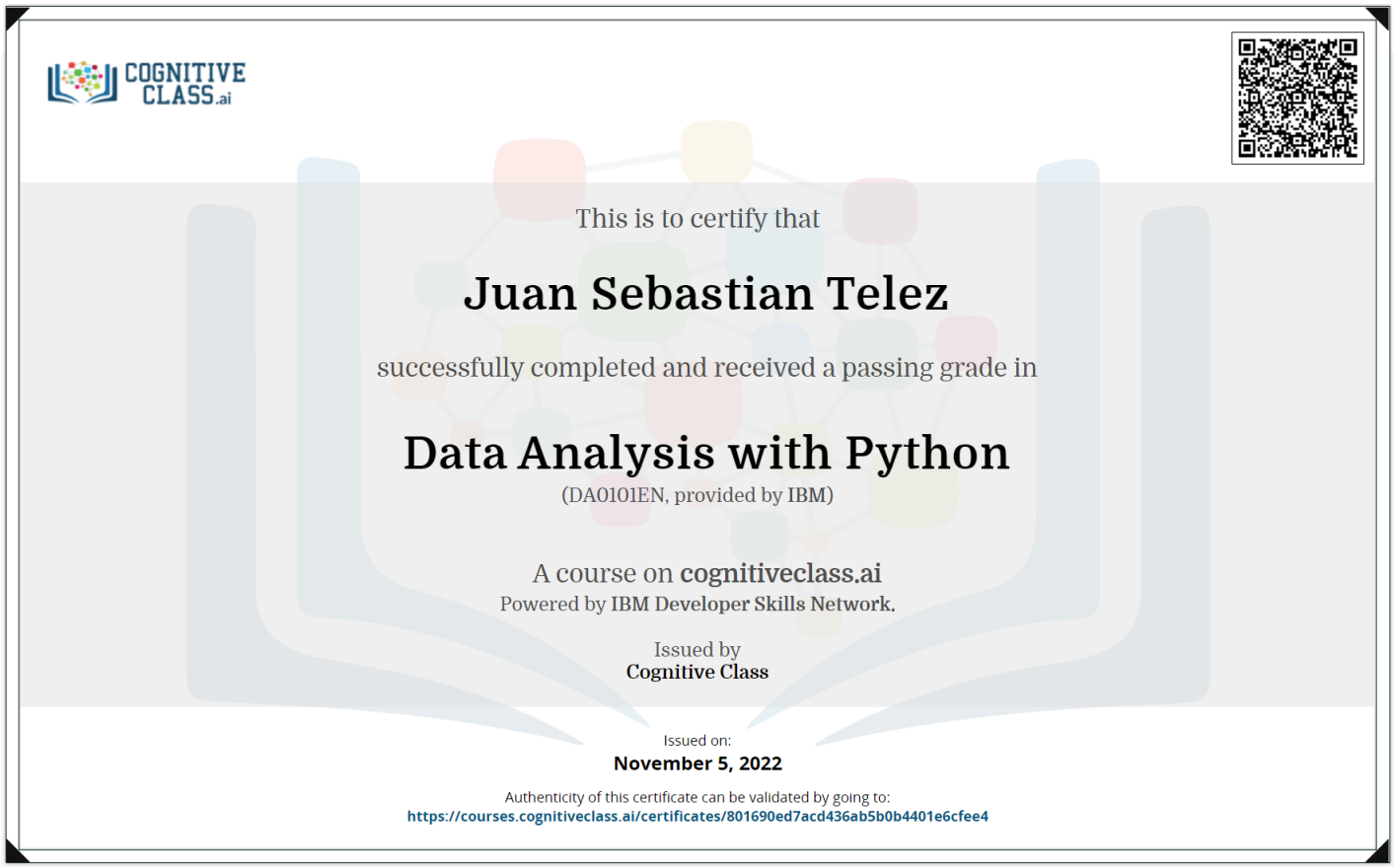
Juan Sebastián Téllez López. A01793859.

Adicionalmente en el archivo "Python by Jtellez" comparto ejercicios realizados por mi cuenta y una guia de aprendizaje de Python.

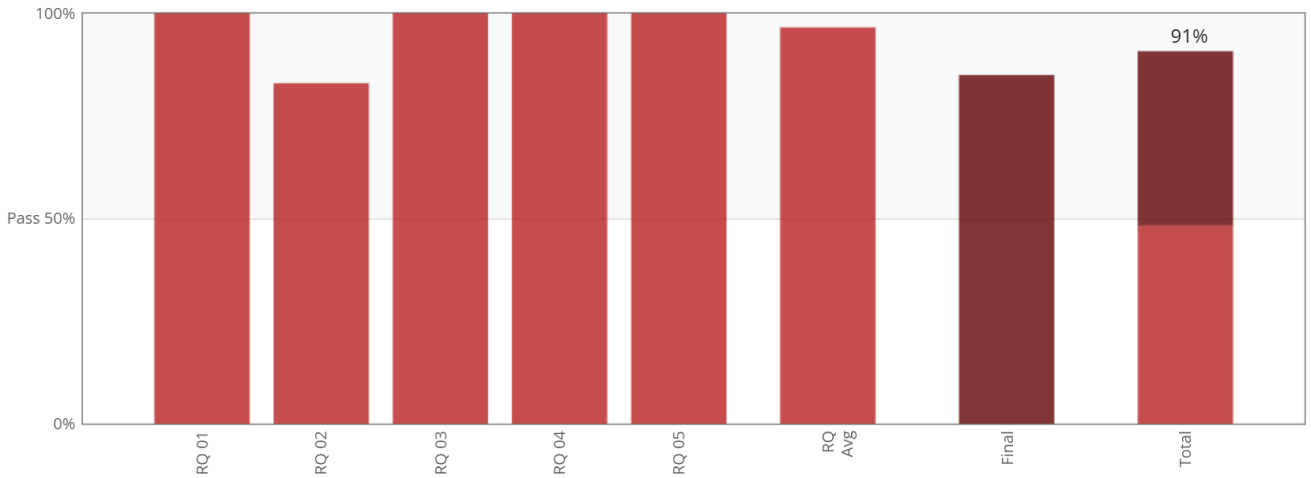


Course Progress for 'Jtellez624' (A01793859@tec.mx)

Congratulations, you qualified for a certificate!

You've earned a certificate for this course.

Request Certificate



Module 1 - Introduction

Introduction to Data Analysis with Python

En este primer modulo se aprendera sobre

- Un problema que requiera analisís de datos.
- Un dataset que será analizado con Python.
- Mirada general a las librerias que podemos utilizar con Python.
- Importar y exportar datos con Python.
- Obtener información reelevante del dataset.

Algunas preguntas que podríamos responder serían.

- Podemos estimar el precio de un vehiculo usado de acuerdo a sus caracteristicas?

The Problem

↳ *1 cell hidden*

Understanding the Data

↳ *4 cells hidden*

Python Packages for Data Science

↳ *3 cells hidden*

Importing and Exporting Data in Python

↳ *2 cells hidden*

Getting Started Analyzing Data in Python

↳ *4 cells hidden*

Lab 1

↳ *3 cells hidden*

Graded Review Questions 1

Question 1

1/1 point (graded)

What does CSV stand for?

- (●) Comma-separated values
- ( ) Car sold values
- ( ) Car state values
- ( ) None of the above

✔

Sav

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 2

1/1 point (graded)

In the data set, which of the following represents an attribute or feature?

- ○ Row
- ◉ Column
- ○ Each element in the dataset

✔

Sav

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 3

1/1 point (graded)

What is the name of what we want to predict?

- (●) Target
- ( ) Feature
- ( ) Dataframe

✔

Sa

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 4

1/1 point (graded)

What is the command to display the first five rows of a dataframe `df`?

- (●) `df.head()`
- ( ) `df.tail()`

✔

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Question 5

1/1 point (graded)

What command do you use to get the data type of each row of the dataframe `df` ?

- (●) `df.dtypes`
- ( ) `df.head()`
- ( ) `df.tail()`

✔

Sav

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 6

1/1 point (graded)

How do you get a statistical summary of a dataframe `df`?

- (●) `df.describe()`
- ( ) `df.head()`
- ( ) `df.tail()`

✔

Sav

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 7

1/1 point (graded)

If you use the method `describe()` without changing any of the arguments, you will get statistical summary of all the columns of type "object".

- (●) False
- ( ) True

✔

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Module 2 - Data Wrangling

↳ *50 cells hidden*

Module 3 - Exploratory Data Analysis

↳ *47 cells hidden*

Module 4 - Model Development

↳ *80 cells hidden*

Module 5 – Model Evaluation

↳ *57 cells hidden*

Final Exam

Question 1

1/1 point (graded)

What does the following command do?

```
df.dropna(subset=["price"], axis=0)
```

- (●) Drop the “not a number” values from the column "price".
- ( ) Drop the row "price".
- ( ) Rename the dataframe "price".

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 2

1/1 point (graded)

How would you provide many of the summary statistics for all the columns in the dataframe "df"?

- (●) df.describe(include = "all")
- ( ) df.head()
- ( ) type(df)
- ( ) df.shape

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 3

1/1 point (graded)

How would you find the shape of the dataframe df?

- ○ df.describe()
- ○ df.head()
- ○ type(df)
- ◉ df.shape

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 4

1/1 point (graded)

What task does the following command, df.to_csv("A.csv"), perform:

- ○ Change the name of the column to "A.csv".
- ○ Load the data from a csv file called "A" into a dataframe.
- ◉ Save the dataframe df to a csv file called "A.csv".

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 5

1/1 point (graded)

What task does the following line of code perform?

```
result = np.linspace(min(df["city-mpg"]), max(df["city-mpg"]), 5)
```

- (●) Builds a bin array ranging from the smallest value to the largest value of "city-mpg" in order to build 4 bins of equal length.
- ( ) Builds a bin array ranging from the smallest value to the largest value of "city-mpg" in order to build 5 bins of equal length.
- ( ) Determines which bin each value of "city-mpg" belongs to.

✔

Show answer

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

Question 6

1/1 point (graded)

What task does the following line of code perform:

```
df['peak-rpm'].replace(np.nan, 5,inplace=True)
```

- (●) Replace the "not a number" values with 5 in the column 'peak-rpm'.
- ( ) Rename the column 'peak-rpm' to 5.
- ( ) Add 5 to the dataframe.

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 7

0/1 point (graded)

How do you "one-hot encode" the column 'fuel-type' in the dataframe df?

- ○ pd.get_dummies(df["fuel-type"])
- ● df.mean([“fuel-type”])
- ○ df[df["fuel-type"])==1]=1

✖

Show answer

Submit You have used 2 of 2 attempts

✖ Incorrect (0/1 point)

Question 8

0/1 point (graded)

What does the vertical axis on a scatterplot represent?

- ○ Independent variable
- ● Dependent variable

Show answer

Submit You have used 1 of 1 attempt

✖ Incorrect (0/1 point)

Question 9

Question 10

1/1 point (graded)

If we have 10 columns and 100 samples, how large is the output of df.corr()?

- ○ 10 x 100
- ◉ 10 x 10
- ○ 100x100
- ○ 100x100

✔

Show answer

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

Question 11

1/1 point (graded)

What is the largest possible element resulting in the following operation "df.corr()"?

- ○ 100
- ○ 1000
- ◉ 1

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 12

1/1 point (graded)

If the Pearson Correlation of two variables is zero:

- ○ The two variable have zero mean.
- ◉ The two variables are not correlated. ✔

Show answer

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Question 13

1/1 point (graded)

If the p-value of the Pearson Correlation is 1:

- ○ The variables are correlated.
- ○ The variables are not correlated.
- ◉ None of the above. ✔

Show answer

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

Question 14

1/1 point (graded)

What does the following line of code do: lm = LinearRegression()?

- ○ Fit a regression object "lm".
- ◉ Create a linear regression object.
- ○ Predict a value.

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 15

1/1 point (graded)

If the predicted function is:

Yhat = a + b1 X1 + b2 X2 + b3 X3 + b4 X4

The method is:

- ○ Polynomial Regression
- ◉ Multiple Linear Regression

✔

Show answer

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Question 16

1/1 point (graded)

What steps do the following lines of code perform:

Input=[('scale',StandardScaler()),('model',LinearRegression())]

pipe=Pipeline(Input)

pipe.fit(Z,y)

ypipe=pipe.predict(Z)

- ○ Standardize the data, then perform a polynomial transform on the features Z.
- ○ Find the correlation between Z and y.
- ◉ Standardize the data, then perform a prediction using a linear regression model using the features Z and targets y.

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 17

1/1 point (graded)

What is the maximum value of R^2 that can be obtained?

- ○ 10
- ◉ 1
- ○ 0

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 18

1/1 point (graded)

We create a polynomial feature PolynomialFeatures(degree=2). What is the order of the polynomial?

- ○ 0
- ○ 1
- ◉ 2

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 19

1/1 point (graded)

You have a linear model. The average R^2 value on your training data is 0.5. You perform a 100th order polynomial transform on your data, then use these values to train another model. Your average R^2 is 0.99. Which comment is correct?

- ○ 100th order polynomial will work better on unseen data.
- ○ You should always use the simplest model.
- ◉ The results on your training data is not the best indicator of how your model performs. You should use your test data to get a beter idea.

✔

Save | Show answer

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 20

0/1 point (graded)

You train a ridge regression model. You get a R^2 of 1 on your validation data and you get a R^2 of 0.5 on your training data. What should you do?

- (●) Nothing. Your model performs flawlessly on your validation data.
- ( ) Your model is under fitting perform a polynomial transform.
- ( ) Your model is overfitting, increase the parameter alpha.

Show answer

Submit You have used 2 of 2 attempts