



Tecnológico de Monterrey

MAESTRÍA EN INTELIGENCIA ARTIFICIAL APLICADA

MATERIA

Ciencia y Analítica de Datos

PROFESOR:

Dr. Jobish Vallikavungal

ALUMNO:

Carlos Enriquez Gorgonio
A01793102

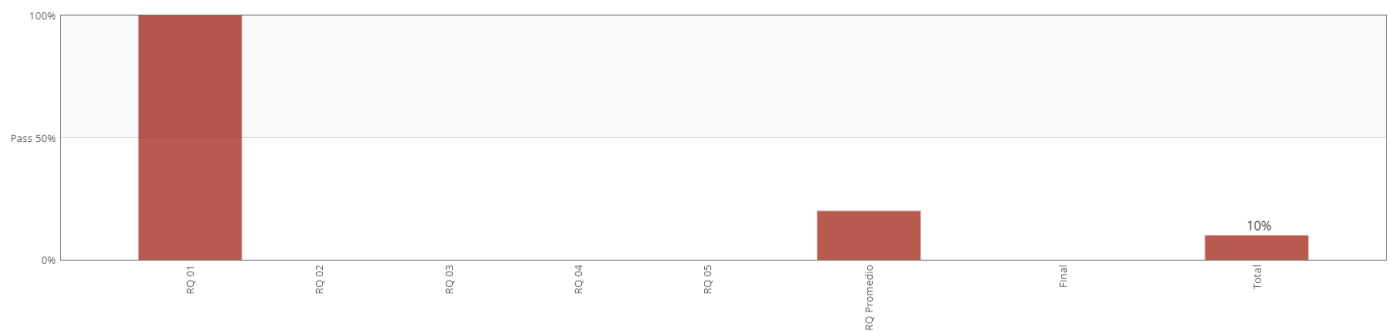
Data Analysis with Python

Octubre de 2022

El vínculo a la actividad en GitHub es:

<https://github.com/PosgradoMNA/actividades-de-aprendizaje-KarltonBotics>

Progreso del curso para el estudiante 'KarltonTec' con el correo (a01793102@tec.mx)



Para leer un archivo csv usamos el método read, de la siguiente forma

```
pandas.read_csv()
```

Para descargar un archivo de internet

```
path = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDDeveloperSkillsNetwork-DA0101EN-SkillsNetwork/labs/Data%20files/auto.csv"
```

```
await download(path, "auto.csv")
```

```
path="auto.csv"
```

```
df = pd.read_csv(path, header=None)
```

Para imprimir los 10 primeros elementos del encabezado de nuestro data set, usamos head, y para los últimos tail

```
print("The last 10 rows of the dataframe\n")  
df.tail(10)
```

Para reemplazar un valor, con un NAN, para que la función dropna puede remover valores faltantes usamos

```
df1=df.replace('?',np.NaN)
```

Para borrar los valores de una columna usamos dropna, en el eje 0, en el eje 1 son filas horizontales

```
df=df1.dropna(subset=["price"], axis=0)
```

Para imprimir las columnas utilizamos:

```
print(df.columns)
```

Para guardar un dataset elaborado en un archivo CSV o coma separated values usamos

```
df.to_csv("automobile.csv", index=False)
```

Para ver el tipo de un dataset

```
print(df.dtypes) o solo df.dtypes
```

Para ver un resumen de datos de un dataset usamos

```
dataframe.describe()
```

Para la información usamos

```
dataframe.info()
```

Para seleccionar las columnas de un data frame las nombramos de la siguiente forma

```
dataframe[['column 1 ',column 2', 'column 3']]
```

Para describir el tipo de cada columna usamos

```
dataframe[['column 1 ',column 2', 'column 3'] ].describe()
```

Para ver el tipo de datos de una columna seleccionada usamos

```
df[['length', 'compression-ratio']].describe()
```