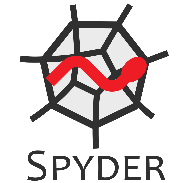
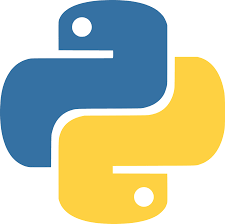
EDEM WORKGROUP  
STATISTICS WITH PYTHON

[Seleccionar fecha]

# ANALYSIS DATA REPORT – SHORT INTERPRETATION

# 



**WORK DONE BY:**

Rocío Jordá

Ruth Gamero

Ander Fdez.

David FC.

Sergio LC

## FROM\_0 INITIAL CLASS

|  |  |
| --- | --- |
|  | [Describa cómo surgió este proyecto, quién participa en él y qué objetivo persigue]  Nota: Para sustituir el texto de cualquier sugerencia (como esta) por el suyo, simplemente haga clic en él y empiece a escribir. Si todavía no quiere agregar ningún texto, simplemente haga clic en la sugerencia y presione la barra espaciadora para borrarlo. |

* Python: object oriented language
* When coding, evitar MAYUSC, evitar caracteres como puntos, comas, %, espacio blanco (usar \_), acentos o palabras largas
* No es 'MI' código – facilitar el trabajo de ‘interpretacion’ al equipo
* String siempre con comillas '…'
* Se crea un INDICE asociado a cada valor de la variable. EMPIEZA SIEMPRE POR CERO. Ojo con las funciones
* Uso de corchetes, parentesis y llaves:

[] = listas - mutables (se pueden cambiar elementos una vez configurada la lista)

() = variables y tuplax - inmutables

{} = diccionarios - mutables

1. Definir variables (lista >> [ ]) --- strings siempre con comillas ''

NAME = ['Ruth', 'Miriam']

AGE = [23, 32]

1. DELETE/CLEAN UP Borrar variables para borrar todas las variables comando reset -f en consola

del (name, age)

1. PRINT imprimir el resultado

Print (age, name)

Print (class\_2019.age)

1. OS.GETCWD get working directory Donde se están guardando las variables? (os library)

os.getcwd()

1. OS.CHDIR change dir Cambiar el directorio donde se estan guardando las variables (os library)

os.chdir('C:/Users/Ruth/Documents/EDEM\_MDA\_2019/02\_CLASES/4\_140919\_Alberto\_Python/Script')

1. PD.DATAFRAME ({XXXX}) Definir Dataframe (panda library) Abrir y ‘observar’ el resultado (DB) en el explorador de variables

class\_2019 = pd.DataFrame ({'name':NAME,'age':AGE}] desde CERO, a partir de variables

rentals\_2011 = pd.read\_csv (xxxxx.csv, sep = ';', decimal = ',') extraccion de CSV (ojo abrir con NOTEPAD ++ para chequear el separador columnas y el separarador decimal y la calidad de los datos)

1. xxx. SHAPE saco filas x columnas (pj. 365 x 11)

class2019.shape

1. xxx. SIZE saco numero total de objetos (pj. 4.015)

class2019.size

1. Xxx.DTYPES define el tipo de variables int64 (numeros sin decimales), object (fecha), float64 (decimales)

rentals\_2011\_ruth.dtypes

1. Xxx.NDIM saco si es uni o multivariable >> da 1 (serie, 1 variable) o 2 (dataframe, +variables)

rentals\_2011\_ruth.ndim

1. Xxx.HEAD () saco las 5 primeras filas

class2019.head()

1. Xxx.TAIL () saco las ultimas 5 filas

Class2019.tail()

1. X.TO\_EXCEL X.TO\_CSV guardar DB en formato XLS o CSV

class\_2019.to\_excel('class\_2019.xlsx')

class\_2019.to\_csv('class\_2019.csv')

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* (esta parte falta por desarrollar\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

VARIABLES NOMINALES (si-no, m-f) o CUANTITATIVAS ORDINALES (bajo-medio-alto, soleado-nublado-lluvioso)

* ESTADISTICO Porcentaje
* GRAFICO Barras

VARIABLES CUANTITATIVAS CONTINUAS (pj. Edad, nivel satisfacion 1-2-3-4-5, temperatura)

ESTADISTICO Media, Desviacion Tipica ('descriptives')

Class\_2019.age.describe()

Count 2

Mean 27 (suma de los elementos dividido por el numero de elementos)

Std

Min

25%

50% (valor del elemento situado en la posición media de la distribucion)

75%

Max

Dtype

GRAFICO Histogram

Plt.hist (class\_2019.age) ( matplotlib.pyplot library)

## DESCRIBE HOLIDAY DATA (Numerically & Graphically)

|  |  |
| --- | --- |
|  | [El alcance del proyecto define los límites de un proyecto. Piense en el alcance como si fuera una caja imaginaria que va a contener todos los elementos o actividades del proyecto. No solo define lo que se hace (lo que entra en la caja), sino que también limita qué no se va a hacer (lo que no cabe en la caja). El alcance siempre responde a preguntas sobre qué se hará, qué no se hará y qué tipo de resultados se obtendrán.] |

(See option 2 for a more complex analysis)

OPTION 1

HOLIDAY analysis (variable included in the Dataframe 2011-2012)

* HOLIDAY includes all Bank Holidays. DOES NOT include weekends (Sundays)
* Numeric Analysis: Frecuency is analized using (PD.CROSSTAB). We get to %Frecuency with ‘Total’ variable

holiday\_table = pd.crosstab(index=wtrent ['holiday'], columns = 'days')

print (holiday\_table)

col\_0

holiday days

0 710

* 1. 21

total = holiday\_table.sum()

holiday\_table\_percent = (holiday\_table/total)\*100

holiday\_table\_percent

col\_0

holiday days

0 97.1%

* 1. 2.9%
* Visual Analysis: Bar chart (PLT.BAR) with additional steps for professional approach

plt.bar(holiday\_table\_percent.index,holiday\_table\_percent['days'])

varx = ('Non-holiday','Holiday')

plt.xticks(holiday\_table\_percent.index, varx)

plt.title('Figure 1.2 2011-2012 % Holidays',loc='center')

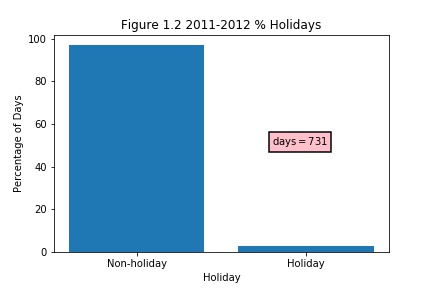
plt.ylabel('Percentage of Days')

plt.xlabel ('Holiday')

legend\_props = dict (boxstyle ='square', facecolor ='pink', lw = 1.5)

textstr = '$\mathrm{days}=%.0f$'%(total)

plt.text (0.8,50, textstr , bbox = legend\_props)



OPTION 2

New variable created to identify not only Bank Holidays but also Weekends. This variable is the difference between two variables from the dataframe: ‘Workingday’ and ‘Holiday’

wtrent ['daytype'] = wtrent ['workingday'] - wtrent ['holiday']

This variable has three different values:

1 = working day

0 = weekend

-1= bank holiday

* Numeric Analysis:

col\_0

daytype days

-1 2.9%

* 1. 28.7%

1 68.4%

* Visual Analysis:

## 

## PROFESSIONAL HISTOGRAM FOR CUSTOMERS

|  |  |
| --- | --- |
|  | [Describa los requisitos a alto nivel del proyecto. Por ejemplo:] |

Es necesario que el nuevo sistema incluya lo siguiente:

* La capacidad de permitir a los usuarios tanto internos como externos que obtengan acceso a la aplicación sin que sea necesario descargar software alguno
* La capacidad para interactuar con la aplicación de almacén de datos existente
* La capacidad para incorporar enrutamiento y notificaciones automatizados en base a reglas empresariales

## BAR CHART SALES BY CASUAL/REGISTERED USERS

|  |  |
| --- | --- |
|  | [Describa las agencias, partes interesadas o departamentos que se verán afectados por este proyecto y explique en qué notarán los efectos.] |

## SALES BY CASUALS/REGISTERED USERS IN WORKING-HOLIDAY DAYS

|  |  |
| --- | --- |
|  | [Describa procesos o sistemas empresariales que se verán afectados por este proyecto y explique en qué notarán los efectos] |

# CHECK LIST

General information about the workgroup..

|  |  |  |
| --- | --- | --- |
| User | In Charge of Point… | Date |
| Rocío Jordá | 2 |  |
| Ruth García |  |  |
| Ander Fdez. |  |  |
| David FC. |  |  |
| Sergio LC |  |  |