TASK 10

Video 1 — Pandas. / /

Pandas is alibaray allows us to ~~read~~ easily read in and work with different type of data.

→ jupiter by default display 20 coloumns

① df = Pandas.read_CSV(' ')
We Put our data in the mentioned file to avariable Called df

df → Print data.
df.Shape → (rows, Coloumns) → returns number of rown and Coloumns.
df.info() → method. return number of rows and Coloums and data about it (Type of data - number of non null Variables ~~etc~~ ---)
pd.set_option('display.max_Columns', 85)
To display all my data
(85 →) عدد الاعمده اللي عايز اعرضه

schema_df = pd.read_CSV('~~dat~~ ---)
To Load The schema CSV → شان احط الـداتا اللي انا بيها كانت الاعمده بتاعتي

df.head(4) → See first 5 rows or Pass a value here

df.tail(4) → See last 5 rows of data or Pass avalue

Video 2: Data frame and selecting series Basics

بديل ال Pandas هو dictionary لو انا بشتغل Pure Python [Keys-Values]
لشخص واحد

لكن لو اكتر من شخص هعمل List of dictionary / dictionary of List
كل شخص هتمثل List

قدر نحول بكل سهوله من dictionary
ل Pandas

Pd = P

df = Pd.Dataframe (People)

→ How to access a single coloumn.
(df['email']) → Type Series or (df.email) → it might gives some errors.

→ What is series object: it is simply a one dimentional array
"rows of data"

→ (~~every column series as~~)
How to access multiple coloumns ??
df [['Last', 'email']

→ df.columns → give me all my columns

→ To get a row (iLoc) integer location:
df.iloc [0]
↑ index of row

→ To get Multiple row
df.iLoc[[0,1]]

→ df.iloc [[0,1], 2] هنا بيبحث بالرقم بتاع index
To access rows and Columns together

df.loc [[0,1]] ← نفس اللي فاتت
To get rows and columns with Loc
we need to pass name of column not index

df.loc[[0,1] , 'email']

→ How to know how many answers of specific answer are?
" df['Hobbyist'] .valueCounts ( )

→ To do slicing
df.loc [0:2 , 'Hobbyist']
↑ inclusive
مش زي
slicing العادي

video 3 : indexes
اصلا بيبقى عندي default index بيبقى كأنه عمود لوحده من غير عنوان
بيبدأ من صفر لكن انا ممكن احدد index اللي انا عايزه

df.set_index(' ↑ ')
name of Coloumn I want to be index
if I want to see my data after this Line it will back to defult : What should I do?
df.set_index('email' , inplace=True)
ده هيسهل عليها اني ابحث في داتا بتاعتي مجبش باقي
الحاجات اللي عايزه

How to resed index ??
df. reset_index(inplace=True)
ممكن اغير الـ index من البداية خالص من وانا بحمل الداتا
pd.read_csv(' ' , index_col='Rest')

How to order My choosen index ALphapiticaly ??

Schema_df.Sort_index() or

Schema_df.Sort_index(ascending=false) الترتيب بالعكس

Video 4: using conditional to filter ~~data~~ rows and columns

> df['Last'] == 'Doe' returns a Series of True and false

Filt = (df['last'] == 'Doe')

df[Filt] → return all dataframe that has last name Doe

= df.loc[Filt]

↳ Last name is Doe and first name is John:

Filt ((df['Last'] == 'Doe') & (df['First'] == 'John'))

ممكن نستخدم

| or ~ opposive.

→ لو انا عايز اعمل Filter فيه اكتر من داله مثلا بدل ما افضل اكتب داله داله ممكن اعمل ليهم List

Countries = ['united States', 'India', 'Germany']

To search for athing in My Survey

Filt = df['LanguageWorkedWith'].str.contains('Python')

Video 5: Updating rows and Columns

→ ex to change all Columns to upper case

```
df.Columns=[x.upper() For x in df.columns]
```

→ remove spaces and put spaces under score

```
df.columns = df.columns.str.replace(' ','_')
```

→ To change name a column.

```
df.rename(columns={'first_name':'firs'}, inplace=True)
```

→ updating data in a row

```
df.loc[2] = ['john', 'smith']
```

or `df.loc[2,['last','email']] = ['Doe', 'john@email']`

or `df.loc[2,'last'] = 'smith'` To change only one value

```
df.at[2,'last'] = 'smith'
```

Filt=(df['email']=='johnDoe@e')

```
df.loc[Filt,'last'] = 'smith'
```

→ update Multiple rows

```
df['email'].str.lower()
```

→ apply function on Series

in a Series `df['email'].apply(len)` ← length of characters

→ apply function on data frame.

`df.apply(len)` → doesn't apply function to each data frame

it only apply for each ~~data frame~~ series

len=3 يعني بتاخد عدد الاحرف في ٣ في

To get min value of data:

```
df.apply(pd.Series.min)
```
→ هيرجع اصغر حاجه في كل عمود

or
```
df.apply(lambda x: x.min())
```

apply على كل عنصر لوحده
applymap مستخدم
كده برده مع نفس تطبيق

```
df.applymap(len)
```
→ returns length of every element

```
df['first'].map({'Corey': 'Chris', 'Jane': 'Mary'})
```
كده هيغير اول اتنين ويخلي التالت
NAN ازاي احل المشكله دي؟

```
df['First'].replace({'Corey': 'Chris', 'Jane': 'Mary'})
```
هتحافظ ع القيمه التالته

Video 6: Add/remove rows and Columns.

How To Combine first and Second Column in one Column.

```
df['first'] + ' ' + df['last']

df['Full_name'] = df['first'] + ' ' + df['Last']
```

→ remove Column

```
df.drop(columns=['First', 'Last']
```

```
df['full-name'].str.split(' ')

df['full_name'].str.split(' ', expand=True)
```

df.describe()
gives me a quick statistical overview
mean - max -

~~add a single row~~

df.append({'first':'Tony','Last':'Mo'}, ignore_index=True)

add data frame to data frame.

df.append(df2, ignore_index=True)

df.append(df2, ignore_index=True)

Video 7: Sorting data

to sort by Last name

df.sort_values(by='Last')

df.sort_values(by='Last', ascending=False)

index للترتيب بـ

df.sort_index()

To Sort Single Column

df['Las'].sort_values()

To get largest of single column

df[' '].nLargest(10)

عدد العناصر اللي انا بريدها

or df.nlargest(10, ' ')

(nSmallest)

---

import matplotlib.pyplot as plt

%matplotlib inline

How to fill missing values

df['i'].fillna(df['i'].mean(), inplace=false)

كما تملئ القيم المفقودة NAN بمتوسط الأرقام

How To getridoff NAN values

df.dropna()