

Republic Yemen

Sana'a of University

College of Computer and IT

Department of Computer Science



الجمهورية اليمنية

جامعة صنعاء

كلية الحاسوب وتكنولوجيا المعلومات

قسم علوم الحاسوب

بسم الله الرحمن الرحيم

عمل الطالب: أسامة محمد صالح علي الهرش

رقم القيد: 21\_16\_4145

المقرر: تنقيب البيانات عملي.

تحت اشراف الدكتورة: نورا الحبابي

التكليف: توثيق خطوات تنظيف Dataset

**الخطوة الأولى: تتكون من عدة نقاط:**

١. تنزيل ملف يحتوي على ال dataset من موقع Kaggle.com.

٢. فتح برنامج ال Jupyter notebook باستخدام لغة البايثون.

٣. انشاء مجلد جديد.

٤. فتح ال cmd تثبيت مكتبة ال pandas.

**الخطوة الثانية: تتكون من عدة نقاط:**

١. عمل `import pandas as pd`

٢. عمل قراءة للملف بشكل التالي والكود يوضح ذلك `df =`

`pd.read_excel('E:\CS4\Term_2\DataMining\labs\Customer Call  
List.xlsx')`

٣. عمل طباعة واعرض للكود السابق.

٤. حذف البيانات او الصفوف المتكررة `(df = df.drop_duplicates`

٥. حذف احد الاعمدة الغير مهمه `df = df.drop(columns = "Not_Useful_Column")`

٦. تغيير بعض ارقام الهواتف من مكان الى اخر  
`df["Phone_Number"] = df["Phone_Number"].str.replace('nan--', '')`

**الخطوة الثالثة: تتكون من عدة نقاط:**

١. إضافة بعض الاعمدة الناقص او البيانات الغير مكتمله مثل  
`df[["Street_Address", "State", "Zip_Code"]] = df["Address"].str.split(',', 2, expand=True)`

٢. المقارنة والتبديل بين عمود واخر مثل  
`df["Do_Not_Contact"] = df["Do_Not_Contact"].str.replace('Yes', 'Y')`

`df["Do_Not_Contact"] = df["Do_Not_Contact"].str.replace('No', 'N')`

٣. حذف الصفوف الفارغة التي لا تحتوي على بيانات  
`df=df.fillna("")`

٤. فرز وتنظيم البيانات على حسب ال index مثل  
`for x in df.index`

`:if df.loc[x, "Do_Not_Contact"] == 'Y'`

`df.drop(x, inplace=True)`

**الخطوة الرابعة والأخيرة :** تم تنظيف البيانات وتقليصها وحذف الأخطاء منها وترتيبها بحسب ال index واطهرها بشكل واضح:

Out[14]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Street_Address	State	Zip_Code
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes	N	123 Shire Lane	Shire	
1	1005	Jon	Snow	876(678)3469	123 Dragons Road	Y	N	123 Dragons Road		
2	1008	Sherlock	Holmes	876(678)3469	98 Clue Drive	N	N	98 Clue Drive		
3	1009	Gandalf		N/a	123 Middle Earth	Yes		123 Middle Earth		
4	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes	N	25th Main Street	New York	
5	1013	Don	Draper	123-543-2345	2039 Main Street	Yes	N	2039 Main Street		
6	1014	Leslie	Knope	876(678)3469	343 City Parkway	Yes	N	343 City Parkway		
7	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	N	N	214 HR Avenue		
8	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	N	2395 Hogwarts Avenue		
9	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes	N	121 Paper Avenue	Pennsylvania	
10	1020	Anakin	Skywalker	876(678)3469	910 Tatooline Road, Tatooline	Yes	N	910 Tatooline Road	Tatooline	
11	1020	Anakin	Skywalker	876(678)3469	910 Tatooline Road, Tatooline	Yes	N	910 Tatooline Road	Tatooline	

In [ ]: