

The background of the slide is a dense field of 3D-rendered numbers in various shades of blue and white. The numbers are of different sizes and are scattered across the entire frame, creating a sense of depth and movement. Some numbers are in the foreground, appearing larger and more detailed, while others are in the background, appearing smaller and more faded. The overall effect is a vibrant, abstract pattern of digits.

The impact of women's driving on “WOSUL” program

Created by:

Eng.Hessah Abdulaziz Alnajem



Agenda:

- Question/need.
- Data Description.
- Tools.
- MVP Goal.

Question/need.

- I will build a model system that , I looking to see "The impact of women's driving on "WOSUL" program"
- The Target users: **Women's in Saudi Arabia.**



Question/need.

Q) What is “WOSOL” program”?

برنامج "وصول" لنقل المرأة العاملة

تحقيقاً لرؤية المملكة العربية السعودية 2030 وبرنامج التحول الوطني 2020 ينطلق برنامج نقل المرأة العاملة "وصول" والذي يهدف لتمكين المرأة من العمل في القطاع الخاص ورفع مشاركتها في سوق العمل عن طريق مساعدة الموظفة في تخطي صعوبات المواصلات من وإلى مكان العمل و ذلك في سبيل دعم استقرارها الوظيفي يعمل البرنامج بدعم مالي من صندوق تنمية الموارد البشرية و بالتعاون مع تطبيقات توجيه المركبات لتقديم خدمة النقل بجودة عالية وبتكلفة مناسبة



To know more about this program, visit the website below:

<https://wusool.sa/service.html>

Data Description:

- Our data is built of 123871 data point and 8 fields . As the link
- It's from <https://data.gov.sa/> open data source.
- Our data sample is **women in Saudi Arabia that can get “WOSUL” program after women driving.**
- I will predict that **the number of women's reduce as the women's start driving.**
- I will model the data set **using classification (Random Forest) Algorithm.**

Data Description:

Step 0: Imports

```
In [60]: #import packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import datetime as dt
from sklearn.ensemble import RandomForestClassifier
from sklearn.datasets import make_classification
```

Step 1: Dataset

```
In [6]: # read the data set
df=pd.read_csv('wusool_data.csv')
```

```
In [7]: # show the 1st two rows in data set
df.head(2)
```

Out[7]:

رقم تسلسلي	الحالة	المدينة	المنطقة	مدينة العنوان الوطني	تاريخ أول رحلة	تاريخ آخر رحلة	تاريخ التسجيل
1 0	فعالة (قامت بأخذ رحلة واحدة على الأقل في البرن...	ثول	مكة المكرمة	NaN	4/15/2021	4/16/2021	4/12/2021
2 1	فعالة (قامت بأخذ رحلة واحدة على الأقل في البرن...	طريف	الرياض	NaN	3/2/2021	4/20/2021	3/1/2021

```
In [8]: # show the informations in data set
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123871 entries, 0 to 123870
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   123871                                non-null      int64
1   123871                                الحالة non-null      object
2   119594                                المدينة non-null      object
3   121306                                المنطقة non-null      object
4   96088 مدينة العنوان الوطني non-null      object
5   100594 تاريخ أول رحلة non-null      object
6   100594 تاريخ آخر رحلة non-null      object
7   123871 تاريخ التسجيل non-null      object
dtypes: int64(1), object(7)
memory usage: 7.6+ MB
```

Data Description:

- As the code below we have 8 columns with different data type:

Column's name	Data type
رقم تسلسلي	int64
الحالة	object
المدينة	object
المنطقة	object
مدينة العنوان الوطني	object
تاريخ أول رحلة	object
تاريخ آخر رحلة	object
تاريخ التسجيل	object

Data Description:

- We will change the columns name to English:

Step 2: Data cleaning & prepressing

1- column:

- rename columns to english.
- remove white space.
- check data type.

2- row:

- check missing date.
- fill missing date with 0.
- check duplicate.

```
In [6]: df.columns
```

```
Out[6]: Index(['رقم تسلسلي', 'الحالة', 'المدينة', 'المنطقة', 'مدينة العنوان الوطني',  
              'تاريخ أول رحلة', 'تاريخ آخر رحلة', 'تاريخ التسجيل'],  
              dtype='object')
```

```
In [7]: df.shape
```

```
Out[7]: (123871, 8)
```

```
In [9]: # rename the columns  
df= df.rename(columns={  
    'رقم تسلسلي': 'serial-number',  
    'الحالة': 'status',  
    'المدينة': 'city',  
    'المنطقة': 'region',  
    'مدينة العنوان الوطني': 'NAS',  
    'تاريخ أول رحلة': 'FirstFlightDate',  
    'تاريخ آخر رحلة': 'LastFlightDate',  
    'تاريخ التسجيل': 'RegistrationDate'})
```

```
In [13]: df.columns
```

```
Out[13]: Index(['serial-number', 'status', 'city', 'region', 'NAS', 'FirstFlightDate',  
               'LastFlightDate', 'RegistrationDate'],  
               dtype='object')
```


Data Description:

- We will change the columns name to English to be more professional:

Column's name		Data type
رقم تسلسلي	serial-number	int64
الحالة	status	object
المدينة	city	object
المنطقة	region	object
مدينة العنوان الوطني	NAS	object
تاريخ أول رحلة	FirstFlightDate	object
تاريخ آخر رحلة	LastFlightDate	object
تاريخ التسجيل	RegistrationDate	object

Data Description:

After importing the libraries, and read data , I will start Clean date:

1- column:

- rename columns to English.
- remove white space.
- check data type.

2- row:

- check missing date.
- fill missing date with 0.
- check duplicate.

Data Description:

```
In [11]: #check data type  
print(df.dtypes)
```

```
serial-number    int64  
status           object  
city             object  
region           object  
NAS              object  
FirstFlightDate  object  
LastFlightDate   object  
RegistrationDate object  
dtype: object
```

```
In [12]: #check missing date.  
df.isna().sum()
```

```
Out[12]: serial-number    0  
status                 0  
city                  4277  
region                2565  
NAS                   27783  
FirstFlightDate       23277  
LastFlightDate        23277  
RegistrationDate       0  
dtype: int64
```

```
In [21]: #fill missing date with 0.  
df = df.fillna(0)  
df.isna().sum()
```

```
Out[21]: serial-number    0  
status                 0  
city                   0  
region                0  
NAS                   0  
FirstFlightDate       0  
LastFlightDate        0  
RegistrationDate       0  
dtype: int64
```

```
In [26]: #check duplicate.  
df.duplicated().sum()
```

```
Out[26]: 0
```

```
In [35]: #print("The Dimension of original DF is: ",df.shape)
```

```
The Dimension of original DF is: (123871, 8)
```

Tools:

- I will use :
 - jupyter environment .
 - python programming language with : (*Numpy, Pandas, MATPLOTLIB, Seaborn, Datetime, RandomForestClassification*) **libraries**

MVP Goals:

- Prepare the environment
- Import all the libraries and dependencies.
- Detect the impact of women's driving on “WOSUL” program based on **status**.
- Detect the of women's driving on “WOSUL” program based on **LastFlightDate**.

The background of the slide is a dense field of 3D-rendered numbers in various shades of blue and white. The numbers are of different sizes and are scattered across the entire frame, creating a sense of depth and movement. Some numbers are in the foreground, appearing larger and more detailed, while others are in the background, appearing smaller and more faded. The overall effect is a vibrant, digital-looking pattern.

Thank you ...

Created by:

Eng.Hessah Abdulaziz Alnajem