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

Created by:

Eng.Hessah Abdulaziz Alnajem

Agenda:

- **Design.**
- **Data.**
- **Algorithm.**
- **Tools.**
- **Communication.**

Design

- This project helps to figure out witch feature has a direct and clear impact on **women's driving on "WOSUL" program based on status.** and predict whether **the number of women's reduce as the women's start driving.**
- The Target users: **Women's in Saudi Arabia.**

Design

- **Questions Goal:**

✓Detect the impact of women's driving on “WOSUL” program **based on status**.

✓Detect the impact of women's driving on “WOSUL” program **based on status** and **Region.**

✓Detect the impact of women's driving on “WOSUL” program **based on status** and **Last Flight Date.**

وزارة العمل والتنمية الاجتماعية
المملكة العربية السعودية
وصول
لك لتعملي

Design

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

وزارة العمل والتنمية الاجتماعية
المملكة العربية السعودية
وصول
لك لتعملي

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

https://wusool.sa/service.html

Data

- 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**.

وزارة العمل
والتنمية الاجتماعية
المملكة العربية السعودية
وصول
لك لتعملي

Data

- Dataset Sample:

رقم تسلسلي	الحالة	المدينة	المنطقة	مدينة العنوان الوطني	تاريخ أول رحلة	تاريخ آخر رحلة	تاريخ التسجيل
1	فعالة	ثول	مكة المكرمة		4/15/2021	4/16/2021	4/12/2021
2	فعالة	طريف	الرياض		3/2/2021	4/20/2021	3/1/2021
3	مستبعدة		الرياض				3/10/2021
4	مؤهلة	الخبر	المنطقة الشرقية				3/1/2021
5	فعالة	الحائر (امارة الرياض)	الرياض		11/4/2020	4/20/2021	11/4/2020
6	مؤهلة	طريف	الرياض				3/20/2021
7	فعالة				12/30/2020	2/18/2021	12/21/2020
8	مستبعدة	المسجد	الرياض		12/27/2020	3/20/2021	12/27/2020
9	فعالة	الخبر	المنطقة الشرقية	الدمام	12/1/2020	4/15/2021	11/30/2020
10	مستبعدة	الرياض	الرياض	الرياض	7/27/2020	8/20/2020	7/27/2020

Data

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

1- column:

- rename columns to English.

- remove white space.

- check data type.

- overwriting FirstFlightDate, LastFlightDate, RegistrationDate after changing format

2- row:

- check missing date.

- fill missing date with NAN.

- check duplicate.

- replace statues data columns with:(مستبعدة"→"InActive" , مؤهلة"→"eligible" , فعالة"→"Active")

- replace each region to English

Algorithm

- Our project is **category classification** , so we choose a *Decian Tree* and *Random Forest algorithm.*

- **Model Evaluation and Selection:**

the entire dataset of **123871 record** was split into **70/30 Train vs. Test**.

- **Below the evaluation of each model:**

Decian Tree

- Accuracy: mean accuracy **0.788 (0.036)**

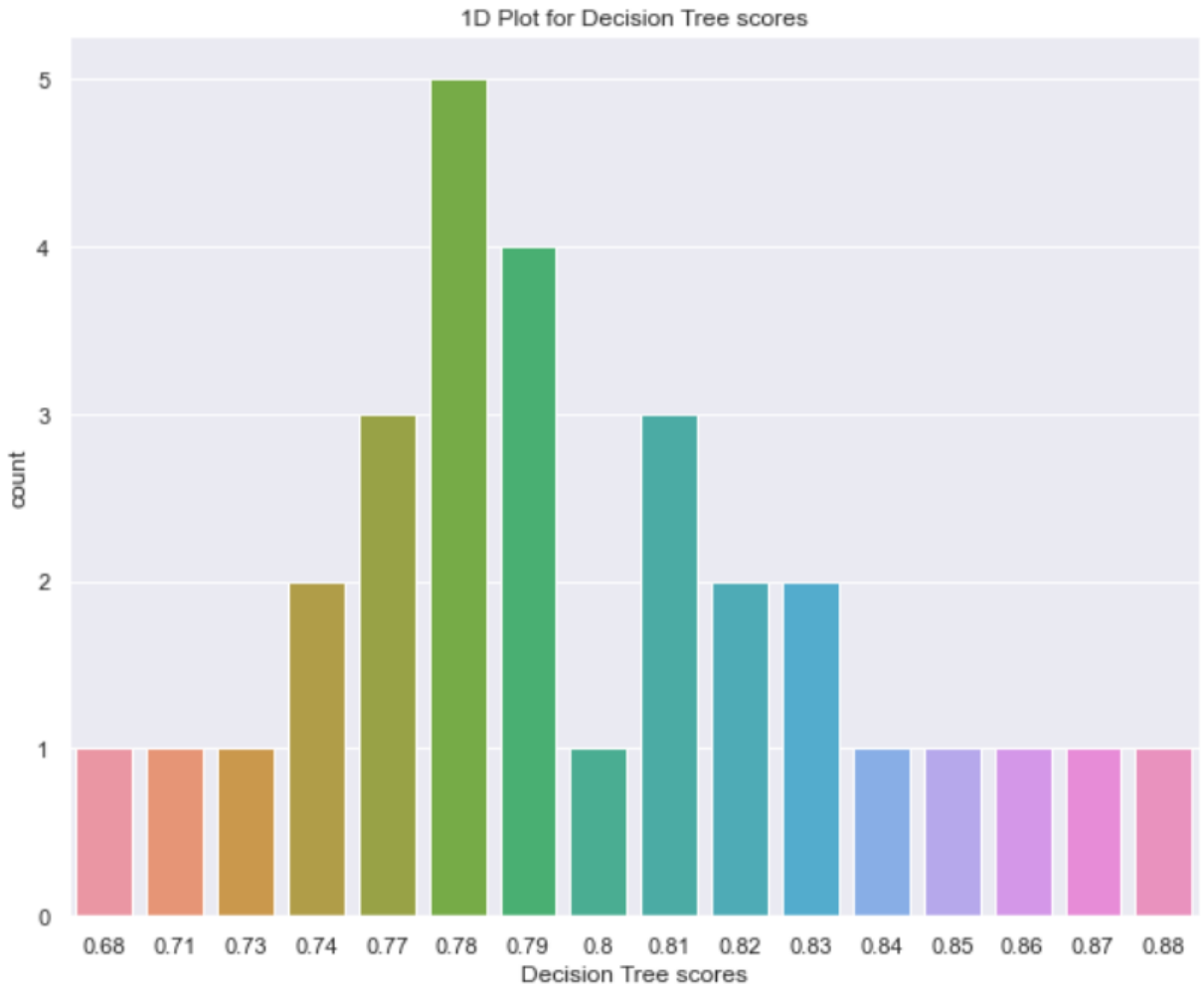
Random Forest

- Accuracy: mean accuracy **0.903 (0.023)**

Communication

1D Plot using countplot:

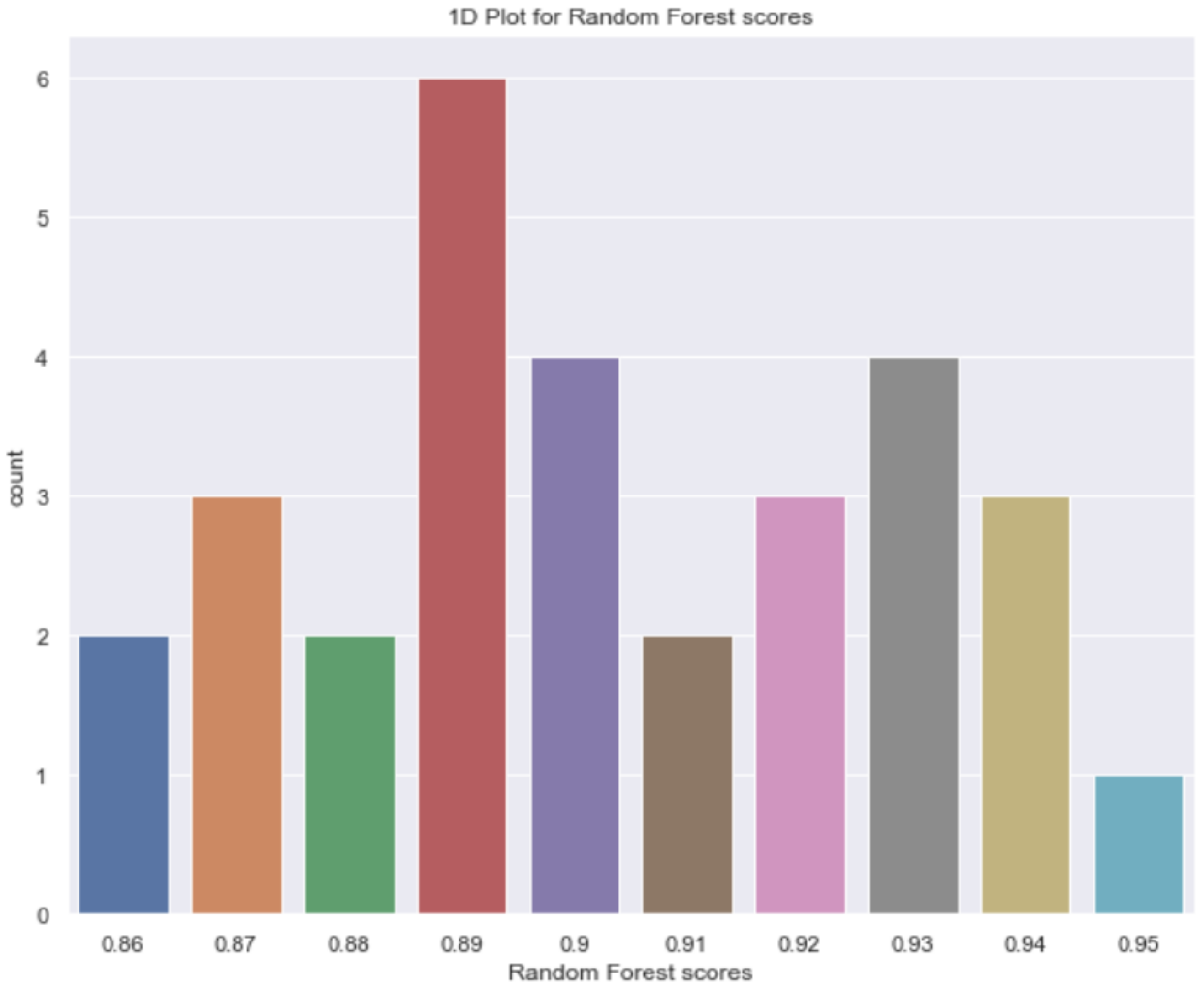
- According to the plot , we can see that ***range number of the n_score*** in the Decian Tree model.
- So, the **mean accuracy** is: **0.788**



Communication

1D Plot using countplot:

- According to the plot , we can see that ***range number of the n_score*** in the Random Forest model.
- So, the mean accuracy is:**0.906**



- This result suggests that women start driving may have a significant negatively impact on "WOSUL" program .

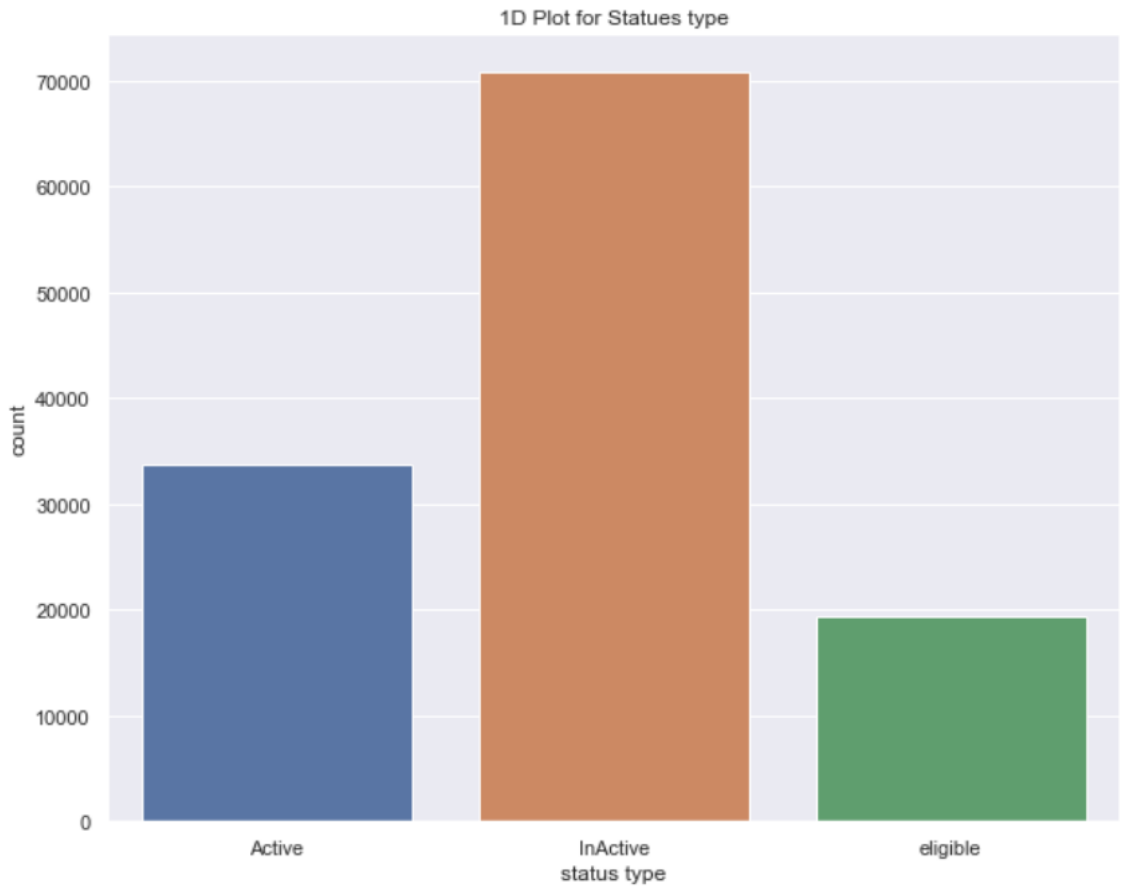
Tools

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

Communication

1D Plot using countplot:

- According to the plot , we can see that number of “InActive” women’s statues is **70000 out of 123871**. So we can it’s the higher number.



Communication

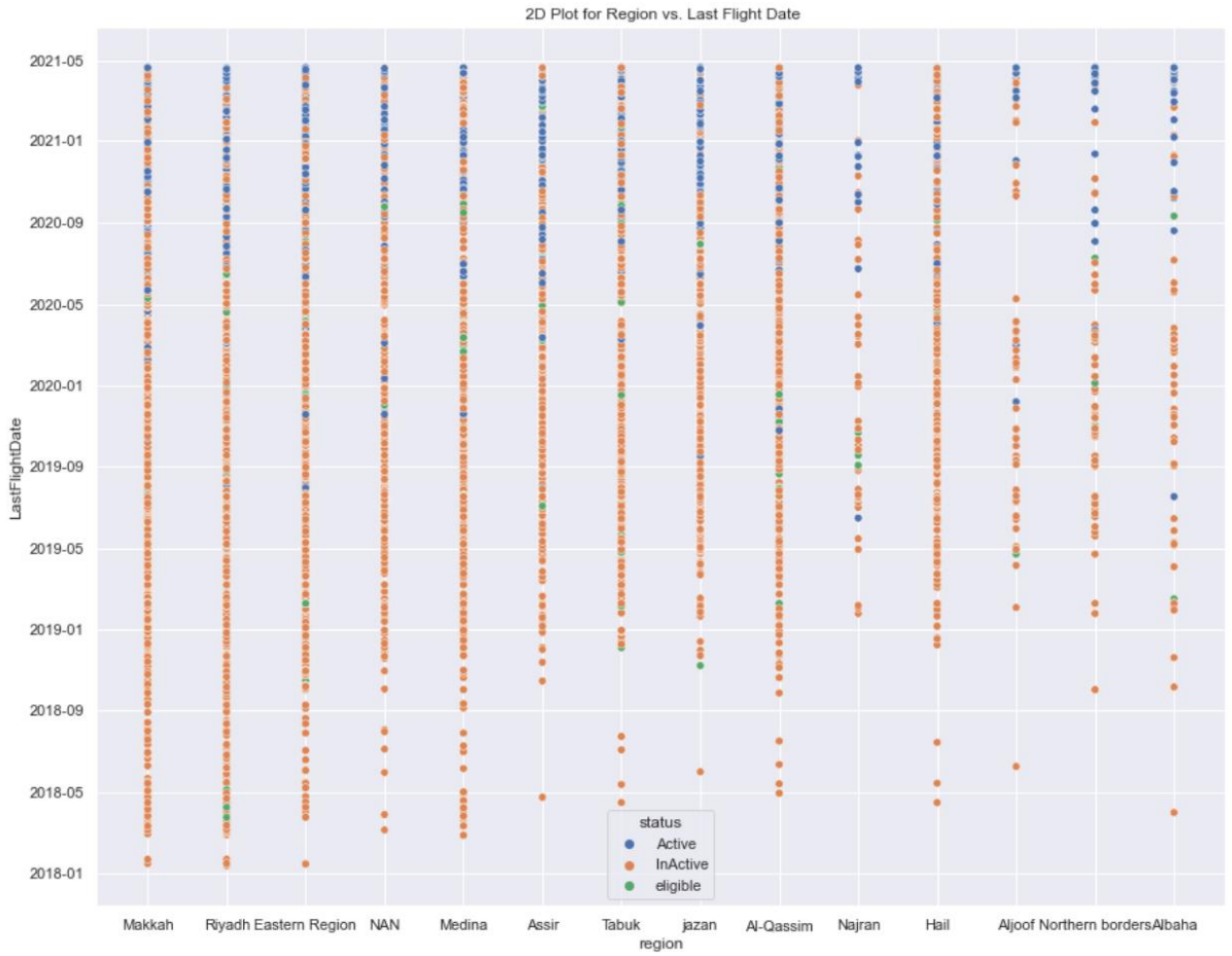
2D Plot using relplot:

- According to the plot , we can see that number of “Abha’s” women’s region is the **lowest value**.
- And “Makkah’s” women’s region is the **highest value**. based on *Region*

Communication

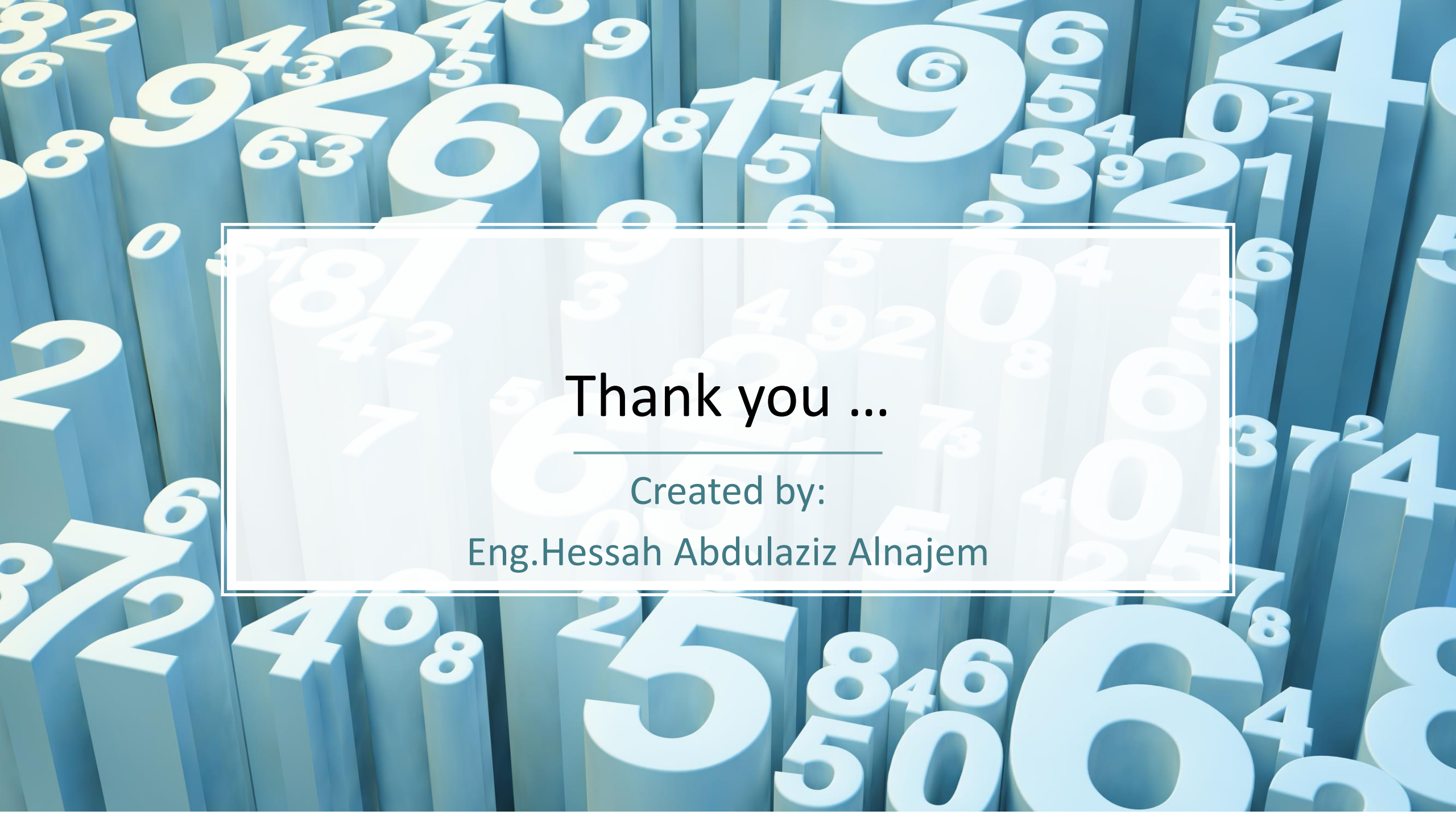
2D Plot using scatterplot :

- According to the plot , we can see that number “InActive” women’s statues the higher number, according to the Last flight date. based on *Last Flight Date*



Finally, we already achieved with the below goals:

- ✓Prepare the environment
- ✓Import all the libraries and dependencies.
- ✓The best model using is **Random Forest** , based on the Accuracy.
- ✓Detect the impact of women's driving on “WOSUL” program **based on status**.
- ✓Detect the impact of women's driving on “WOSUL” program **based on status** and **Region.**
- ✓Detect the impact of women's driving on “WOSUL” program **based on status** and **Last Flight Date.**



Thank you …

Created by:

Eng.Hessah Abdulaziz Alnajem