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

Created by:

Eng.Hessah Abdulaziz Alnajem

Agenda:

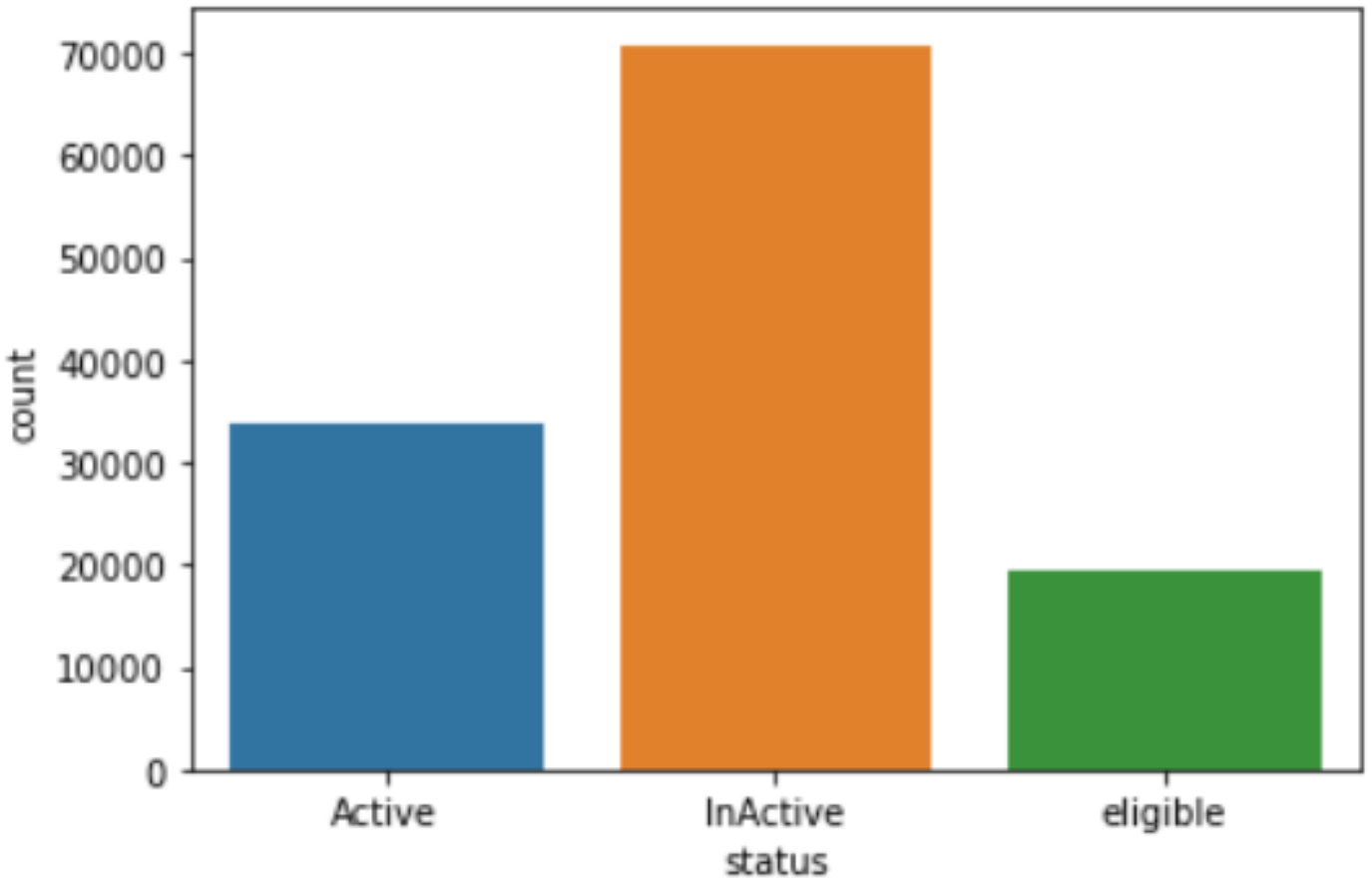
- **Analysis of the project.**
- **MVP Goal.**

Analysis of the project:

- The goal of this project is to better understand impact of women's driving on “WOSUL” program
- To start exploring this goal, I used **a Random Forest model** with one feature/target “stutues” to describe impact of women's driving on “WOSUL” program as a function of the status of “WOSUL” program .

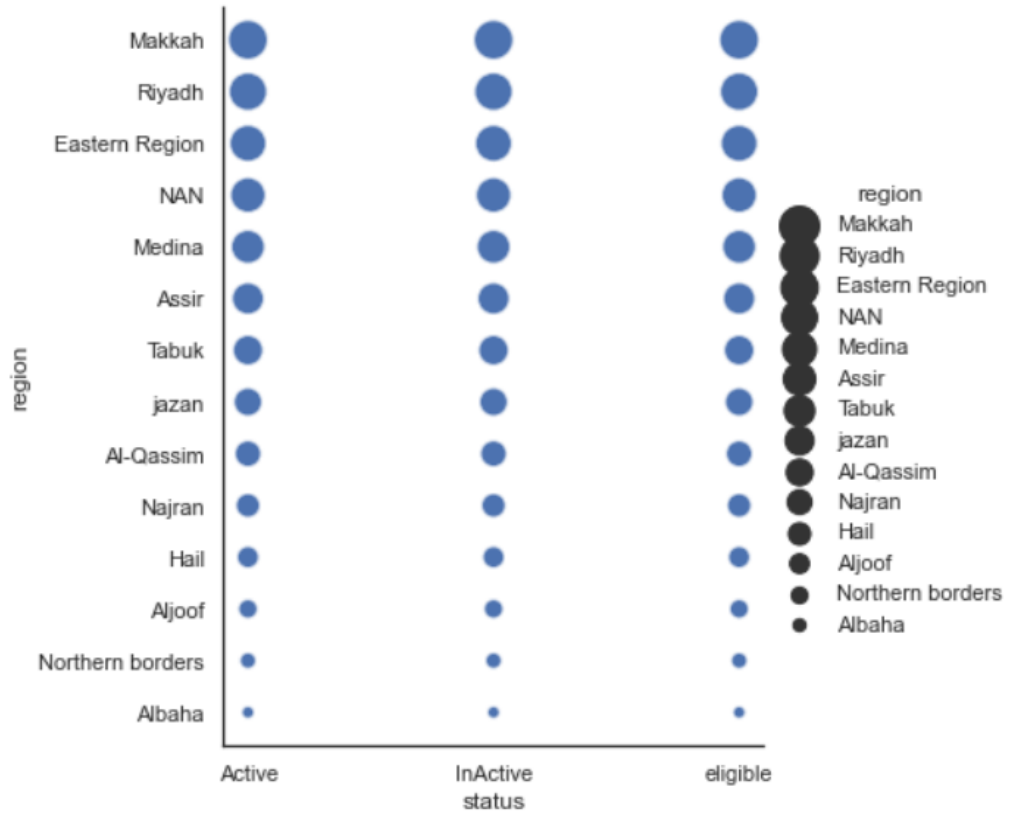
1D Plot using countplot:

- 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.



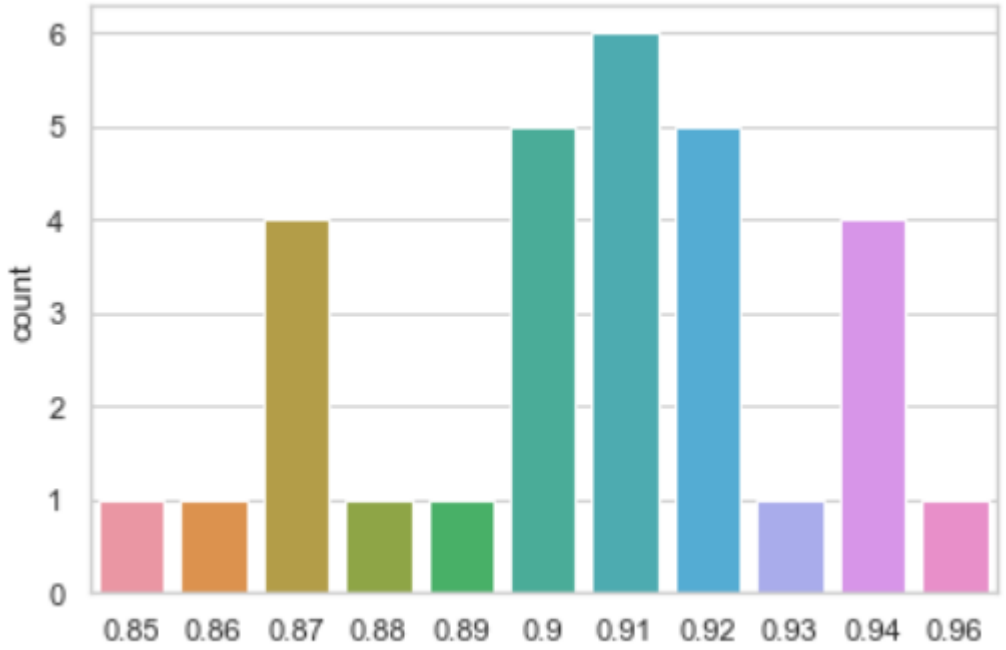
2D Plot using relplot:

- 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**.



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 .

MVP Goals:

Finally, we already achieved with the below 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.**

MVP Goals:

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

Data before cleaning:

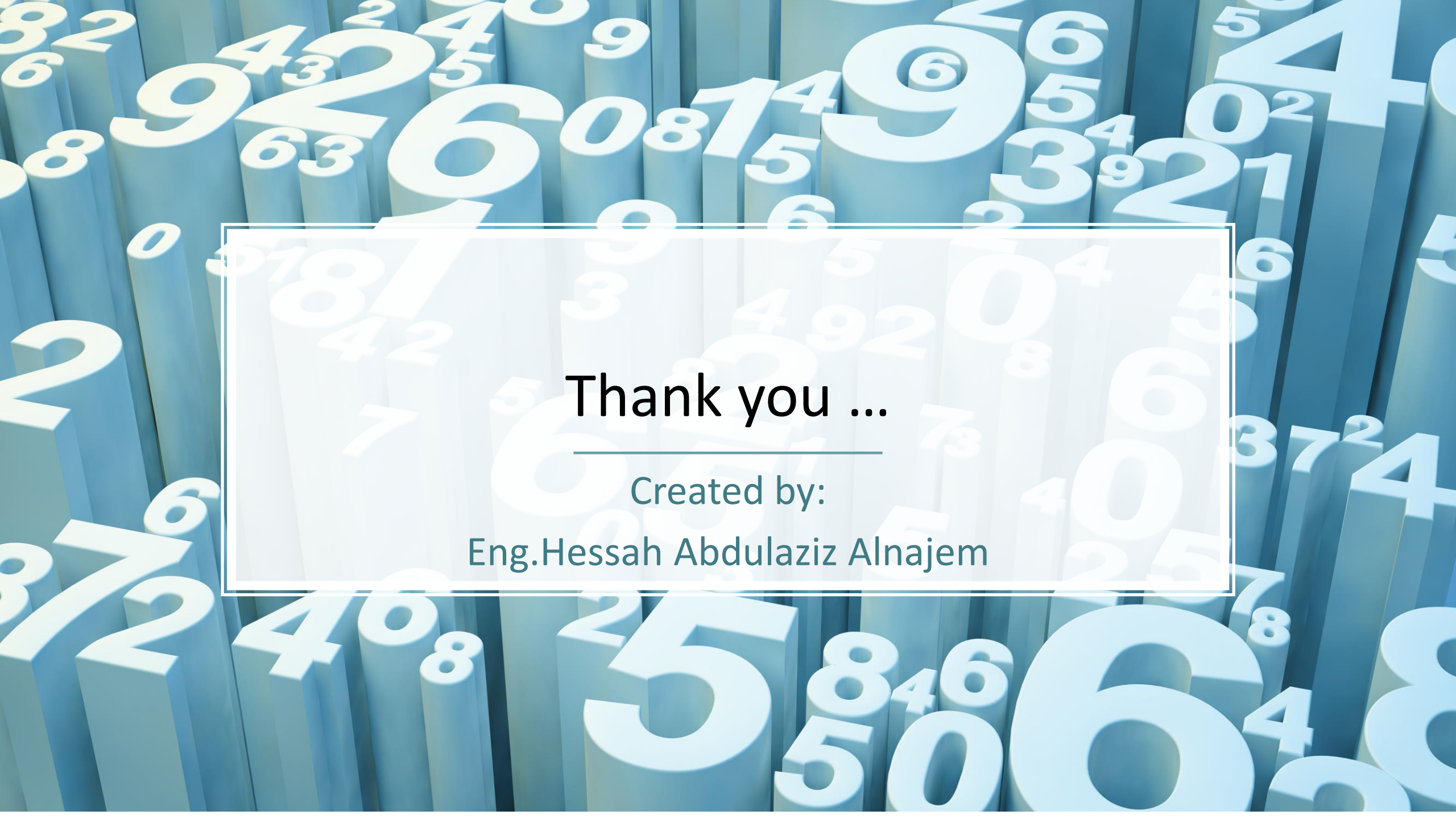
	رقم تسلسلي	الحالة	المدينة	المنطقة	مدينة العنوان الوطني	تاريخ أول رحلة	تاريخ آخر رحلة	تاريخ التسجيل
0	1	فعالة	تول	مكة المكرمة	NaN	4/15/2021	4/16/2021	4/12/2021
1	2	فعالة	طريف	الرياض	NaN	3/2/2021	4/20/2021	3/1/2021
2	3	مستبعدة	NaN	الرياض	NaN	NaN	NaN	3/10/2021
3	4	مؤهلة	الخبر	المنطقة الشرقية	NaN	NaN	NaN	3/1/2021
4	5	فعالة	الحائر (امارة الرياض)	الرياض	NaN	11/4/2020	4/20/2021	11/4/2020

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123871 entries, 0 to 123870
Data columns (total 8 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   123871                  رقم تسلسلي 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 After cleaning:

	serial-number	status	city	region	NAS	FirstFlightDate	LastFlightDate	RegistrationDate
0	1	Active	تول	Makkah	NAN	2021-04-15	2021-04-16	2021-04-12
1	2	Active	طريف	Riyadh	NAN	2021-03-02	2021-04-20	2021-03-01
2	3	InActive	NAN	Riyadh	NAN	NaT	NaT	2021-03-10
3	4	eligible	الخبر	Eastern Region	NAN	NaT	NaT	2021-03-01
4	5	Active	الحائر (امارة الرياض)	Riyadh	NAN	2020-11-04	2021-04-20	2020-11-04

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123871 entries, 0 to 123870
Data columns (total 8 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   serial-number     123871 non-null  int64
 1   status            123871 non-null  object
 2   city              119594 non-null  object
 3   region            121306 non-null  object
 4   NAS               96088 non-null   object
 5   FirstFlightDate   100594 non-null  datetime64[ns]
 6   LastFlightDate    100594 non-null  datetime64[ns]
 7   RegistrationDate  123871 non-null  datetime64[ns]
dtypes: datetime64[ns](3), int64(1), object(4)
memory usage: 7.6+ MB
```



Thank you …

Created by:

Eng.Hessah Abdulaziz Alnajem