King Saud University

College of Business Administration

Management Information Systems Department

COURSE: DSS (MIS350)

2nd Semester 2024

Final Report

Section	NAME	ID
66973	*Rand Aljathlani*	
66973	*Fulwah Aldhuwayhi*	
66973	*Sadeem Almushham*	
66973	*Alhanouf Alhayan*	
66973	*Leena Alateeq*	

Supervised By: Dr. Ramah Al Balawi.

Table of Contents

Phase#1 Project Proposal 3
Project overview: 3
Problem Statement: 3
Link to our datasets: 3
Phase#2: Identify the Key Performance Indicators (KPIs) 4
Our Key Performance Indicators (KPIs): 4
For the Employment Dataset: 4
For the Tourism Dataset: 4
Phase#3: Data Modeling 5
Star Schema for Each Dataset. 5
For The Employment Dataset: 5
For The Tourism Dataset: 6
Phase#4: ETL 7
Phase#5: Data Visualizations 14
Here we'll show the visualization for each KPI, we used tableau to create them. 14
For the Tourism Dataset KPIs: 14
For the Employment Dataset KPIs: 16
Dashboard for Each Dataset: 19
- Tourism Dashboard 19
- Employment Dashboard 20

Conclusion: 20
Work Distribution: 21

Phase#1 Project Proposal

Project overview:

This project comprises several phases. The current report focuses on phase one, which is dedicated to problem formulation. This involves selecting one or more datasets; we chose datasets related to tourism and employment in the Kingdom of Saudi Arabia for the years 2021-2022.

Problem Statement:

The objective of this study is to analyze and understand the patterns and trends associated with tourism in the Kingdom of Saudi Arabia during the years 2021-2022. Additionally, the study aims to examine the dynamics of employment in the Kingdom of Saudi Arabia over the same period. By investigating these aspects, we seek to identify key factors influencing tourism and employment and provide insights that can inform decision-making and policy formulation in the relevant sectors.

Link to our datasets:

- https://studentksuedu-my.sharepoint.com/personal/443200371_student_ksu_edu_sa/Documents/DSS_DATASET.xlsx
- DSS_DATASET.xlsx

Note: Please be advised that both links are identical, yet I am presenting both as a precaution in case one malfunctions.

Phase#2: Identify the Key Performance Indicators (KPIs)

Our Key Performance Indicators (KPIs):

For the Employment Dataset:

1. Number of Employments per Quarter.
2. Number of Employments by Citizenship.
3. Number of Employments per Gender.
4. Ratio of Gender by Citizenship.
5. Ratio of Saudi to Non-Saudi Employment.

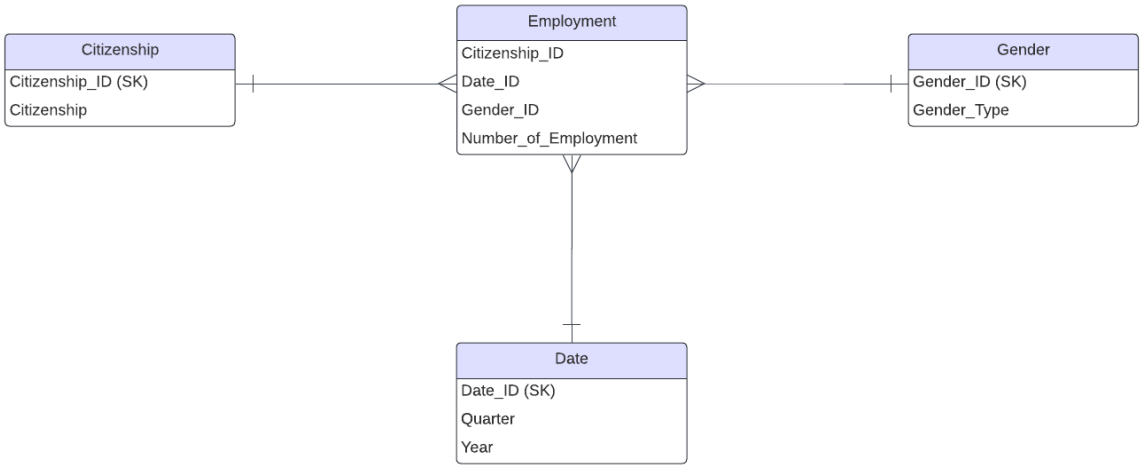
For the Tourism Dataset:

1. Number of Tourists per Month.
2. Number of Tourists by Citizenship.
3. Number of Tourists by Purpose of Trips.
4. Number of Tourists by Mode of Departure.
5. Number of Tourists per Region.

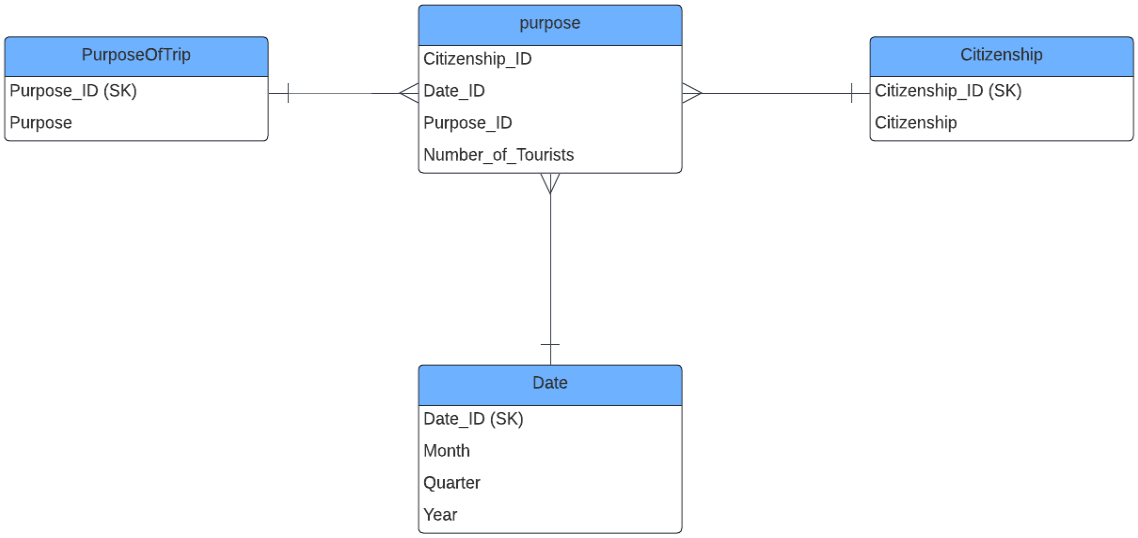
Phase#3: Data Modeling

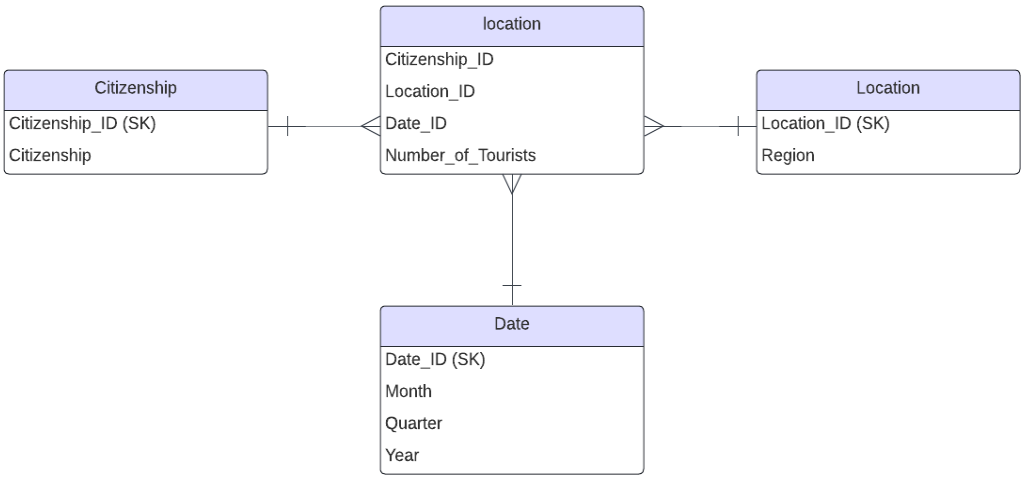
Star Schema for Each Dataset.

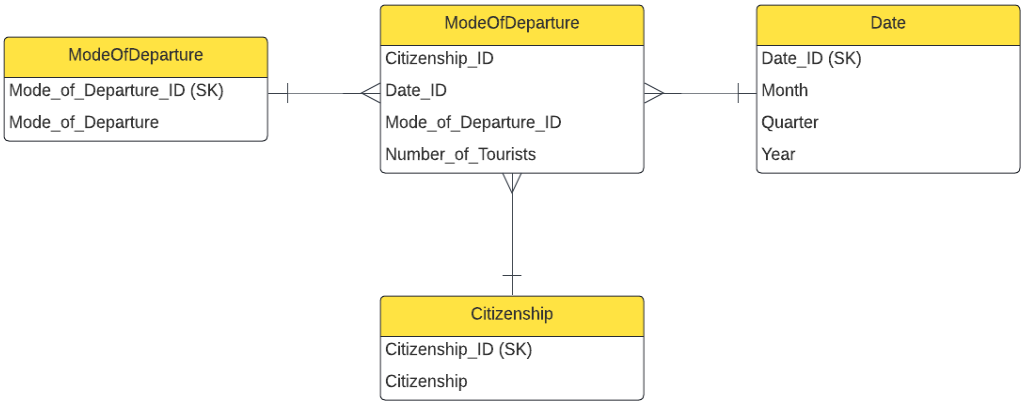
For The Employment Dataset:



For The Tourism Dataset:







Phase#4: ETL

We developed the ETL (Extract, Transform, Load) process using Python and the Pandas library, executed in a Jupyter Notebook environment. We began with extracting datasets pertaining to tourism and employment in the Kingdom of Saudi Arabia for the years 2021-2022 from Saudi open data. These datasets required several transformation steps to ensure they were in a suitable format for analysis and visualization. We translated values from Arabic to English to standardize the dataset and adjusted number formats to ensure numerical data was correctly interpreted and manipulated. Additionally, we converted the data from wide format to long format, simplifying the creation of our fact tables. We also performed data cleaning tasks, such as correcting inconsistencies in categorical data, removing duplicate records, and standardizing text formats, including the names of regions and other categorical variables, to maintain uniformity across the datasets.
Following these transformation steps, we merged relevant tables to form our fact tables, ensuring that each fact table was accurately linked to its corresponding dimension tables. The final step involved loading the transformed data into our data mart, structured according to the star schema we designed. This setup ensured optimized performance for querying and reporting. We then integrated the data mart with Tableau, where we created visualizations to represent our KPIs. This integration allowed for interactive and dynamic reporting, enhancing our ability to derive insights and support strategic decision-making.

In the upcoming sections, you'll find screenshots and explanations of crucial code snippets. These screenshots represent the primary steps in the process, some of which have been repeated. As a result, repetitive actions are not included to keep the focus on essential workflow aspects.

- **Date Dimension**

[1]:
```
import pandas as pd

folder_path = 'DSS_DATASET_UPDATED.xlsx'
df = pd.read_excel(folder_path, sheet_name = 'O2')
# O2 = Purpose of Trips
# O3 =  Number by Destination – Region
# O5 = Mode of Departure

df
```

[1]:

	(Last update: 26-06-2023)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnar
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	O2	Outbound Tourists Number by Main Purpose of Tr...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	Month	Business\nالأعمال	NaN	NaN	Leisure*\nالترفيه*	NaN	NaN	Other\nأخرى	NaN	NaN	F Relativ أو
4	NaN	NaN	NaN	Saudi\nسعودي	Non-Saudi\nغير سعودي	Total\nالإجمالي	Saudi\nسعودي	Non-Saudi\nغير سعودي	Total\nالإجمالي	Saudi\nسعودي	Non-Saudi\nغير سعودي	Total\nالإجمالي	Saud
5	NaN	2021	January	4.912	28.888	33.8	19.87	0	19.87	0	0	0	

Define the file path for the Tourism dataset, then read the sheet ‘O2’

```
[2]: col_list = ['Unnamed: 1', 'Unnamed: 2']

dim_table = df[col_list]

dim_table
```

[2]:

	Unnamed: 1	Unnamed: 2
0	NaN	NaN
1	Outbound Tourists Number by Main Purpose of Tr...	NaN
2	NaN	NaN
3	NaN	Month
4	NaN	NaN
5	2021	January

Create DataFrame ‘dim_table’ with the selected columns.

```
[9]: quarters = []
for month in dim_table['Month']:
    if month in months[0:3]:
        quarters.append('Q1')
    elif month in months[3:6]:
        quarters.append('Q2')
    elif month in months[6:9]:
        quarters.append('Q3')
    elif month in months[9:]:
        quarters.append('Q4')
    else:
        quarters.append(None)

dim_table['Quarter'] = quarters

dim_table
```

[9]:

	Year	Month	Quarter
0	2021	January	Q1
1	2021	February	Q1
2	2021	March	Q1
3	2021	April	Q2
4	2021	May	Q2
5	2021	June	Q2
6	2021	July	Q3
7	2021	August	Q3
8	2021	September	Q3
9	2021	October	Q4
10	2021	November	Q4
11	2021	December	Q4
12	2022	January	Q1
13	2022	February	Q1
14	2022	March	Q1
15	2022	April	Q2
16	2022	May	Q2
17	2022	June	Q2
18	2022	July	Q3
19	2022	August	Q3
20	2022	September	Q3
21	2022	October	Q4

In this cell we create ‘Quarter’ column to ensure consistency between the Emploment Dataset and the Tourism Dataset, because the Date Dimension serves as a shared dimension.

- **Gender Dimension**

```
[5]: gender_mapping = {
    'الذكور': 'Male',
    'الاناث': 'Female'
}
gender_dim_df.loc[:, 'Gender'] = gender_dim_df['Gender'].map(gender_mapping)

gender_dim_df
```

[5]:

	Gender
0	NaN
1	Male
2	Female
3	NaN
4	Male
5	Female
6	NaN
7	Male
8	Female
9	NaN

This part creates a dictionary (gender_mapping) to translate Arabic gender values to English. The mapped values are applied to the 'Gender' column of the DataFrame ‘gender_dim_df’.

```
[6]: gender_dim_df = gender_dim_df.drop_duplicates().dropna().reset_index(drop=True)

     gender_dim_df
```

[6]:

	Gender
0	Male
1	Female

```
[7]: gender_dim_df['Gender_ID'] = range(1, len(gender_dim_df) + 1)
     gender_dim_df
```

[7]:

	Gender	Gender_ID
0	Male	1
1	Female	2

```
[8]: gender_dim_df = gender_dim_df[['Gender_ID', 'Gender']]
     gender_dim_df
```

[8]:

	Gender_ID	Gender
0	1	Male
1	2	Female

This step involves cleaning the DataFrame by removing duplicate records and rows with missing values. The code assigns unique IDs to each gender category and creates a new column 'Gender_ID', essentially acting as the primary key for gender categories.

- **Citizenship Dimension**

```
[5]: citizens_dim_df = citizens_dim_df.drop_duplicates().dropna().reset_index(drop=True)

     citizens_dim_df
```

[5]:

	Citizenship
0	Saudi\n\nسعودي
1	Non-Saudi\n\nغير سعودي
2	Total\n\nالإجمالي

```
[6]: citizens_dim_df['Citizenship'] = citizens_dim_df['Citizenship'].str.split('\n').str[0]

     citizens_dim_df
```

[6]:

	Citizenship
0	Saudi
1	Non-Saudi
2	Total

```
[7]: citizens_dim_df = citizens_dim_df.drop(2)

     citizens_dim_df
```

[7]:

	Citizenship
0	Saudi
1	Non-Saudi

```
[8]: citizens_dim_df['Citizenship_ID'] = range(1, len(citizens_dim_df) + 1)

     citizens_dim_df
```

[8]:

	Citizenship	Citizenship_ID
0	Saudi	1
1	Non-Saudi	2

We remove duplicate rows, splits the values in the "Citizenship" column by newline ('\n') and keeps only the first line of each value, add the ‘Citizenship_ID’ column.

- **Tourism Location Dimension**

```
[9]: def format_region(region):
         words = region.split()
         formatted_words = []
         for word in words:
             if len(word) <= 3:
                 formatted_word = word
             else:
                 formatted_word = word.capitalize()
             formatted_words.append(formatted_word)
         return ' '.join(formatted_words)

     dim_table['Region'] = dim_table['Region'].apply(format_region)

     dim_table
```

[9]:

	Region
0	GCC
1	Middle East
2	Africa
3	Asia
4	Europe
5	America
6	Other Countries

Define a function called (format_region) that takes a region as input, splits it into words, capitalizes each word if it's longer than 3 characters, and then joins the words back into a single string. Then applies this function to the 'Region' column.

- **Tourism Purpose of Trips Dimension**

```
[7]: dim_table['Purpose'] = dim_table['Purpose'].str.replace('*', '')

     dim_table
```

[7]:

	Purpose
1	Business\nالأعمال
2	Leisure\nالترفيه
3	Other\nأخرى
4	Visiting Friends& Relatives\nزيارة الأصدقاء أو...

```
[8]: dim_table['Purpose_ID'] = range(1, len(dim_table) + 1)

     dim_table
```

[8]:

	Purpose	Purpose_ID
1	Business\nالأعمال	1
2	Leisure\nالترفيه	2
3	Other\nأخرى	3
4	Visiting Friends& Relatives\nزيارة الأصدقاء أو...	4

```
[9]: dim_table = dim_table[['Purpose_ID', 'Purpose']]

     dim_table
```

[9]:

	Purpose_ID	Purpose
1	1	Business\nالأعمال
2	2	Leisure\nالترفيه
3	3	Other\nأخرى
4	4	Visiting Friends& Relatives\nزيارة الأصدقاء أو...

```
[10]: dim_table['Purpose'] = dim_table['Purpose'].str.split('\n').str[0]

      dim_table
```

[10]:

	Purpose_ID	Purpose
1	1	Business
2	2	Leisure
3	3	Other
4	4	Visiting Friends& Relatives

Here we used the replace() function to clean the 'Purpose' column, we remove any asterisks found in the values of the 'Purpose' column and replaces them with an empty string.

- **Employment Fact**

[8]:

	Year - Quarter	Saudi Male	Saudi Female	Non-Saudi Male	Non-Saudi Female
0	2021 الربع الأول	1,365,654	723,789	6,051,404	250,388
1	2021 الربع الثاني	1,385,268	680,070	5,869,394	255,438
2	2021 الربع الثالث	1,416,888	718,420	5,762,323	260,754
3	2021 الربع الرابع	1,469,850	770,962	6,010,505	279,991
4	2022 الربع الأول	1,531,720	841,770	6,424,480	298,509
5	2022 الربع الثاني	1,563,771	879,182	6,787,008	311,661
6	2022 الربع الثالث	1,582,946	926,180	6,955,296	318,392
7	2022 الربع الرابع	1,611,085	970,330	7,019,759	321,864

[9]:
```
fact_table = pd.melt(df, id_vars=['Year - Quarter'], var_name="Citizenship and Gender", value_name="Number_of_Employment")

fact_table
```

[9]:

	Year - Quarter	Citizenship and Gender	Number_of_Employment
0	2021 الربع الأول	Saudi Male	1,365,654
1	2021 الربع الثاني	Saudi Male	1,385,268
2	2021 الربع الثالث	Saudi Male	1,416,888
3	2021 الربع الرابع	Saudi Male	1,469,850
4	2022 الربع الأول	Saudi Male	1,531,720
5	2022 الربع الثاني	Saudi Male	1,563,771
6	2022 الربع الثالث	Saudi Male	1,582,946
7	2022 الربع الرابع	Saudi Male	1,611,085
8	2021 الربع الأول	Saudi Female	723,789
9	2021 الربع الثاني	Saudi Female	680,070
10	2021 الربع الثالث	Saudi Female	718,420
11	2021 الربع الرابع	Saudi Female	770,962
12	2022 الربع الأول	Saudi Female	841,770
13	2022 الربع الثاني	Saudi Female	879,182
14	2022 الربع الثالث	Saudi Female	926,180
15	2022 الربع الرابع	Saudi Female	970,330
16	2021 الربع الأول	Non-Saudi Male	6,051,404
17	2021 الربع الثاني	Non-Saudi Male	5,869,394
18	2021 الربع الثالث	Non-Saudi Male	5,762,323
19	2021 الربع الرابع	Non-Saudi Male	6,010,505

In this cell we use the melt() function to transform the dataframe from wide to long.

[10]:
```
fact_table['Citizenship'] = fact_table['Citizenship and Gender'].str.split(' ').str[0]
fact_table['Gender'] = fact_table['Citizenship and Gender'].str.split(' ').str[1]

fact_table.drop(columns=['Citizenship and Gender'], inplace=True)

fact_table
```

[10]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender
0	2021 الربع الأول	1,365,654	Saudi	Male
1	2021 الربع الثاني	1,385,268	Saudi	Male
2	2021 الربع الثالث	1,416,888	Saudi	Male
3	2021 الربع الرابع	1,469,850	Saudi	Male
4	2022 الربع الأول	1,531,720	Saudi	Male
5	2022 الربع الثاني	1,563,771	Saudi	Male
6	2022 الربع الثالث	1,582,946	Saudi	Male
7	2022 الربع الرابع	1,611,085	Saudi	Male
8	2021 الربع الأول	723,789	Saudi	Female
9	2021 الربع الثاني	680,070	Saudi	Female
10	2021 الربع الثالث	718,420	Saudi	Female
11	2021 الربع الرابع	770,962	Saudi	Female
12	2022 الربع الأول	841,770	Saudi	Female
13	2022 الربع الثاني	879,182	Saudi	Female
14	2022 الربع الثالث	926,180	Saudi	Female
15	2022 الربع الرابع	970,330	Saudi	Female
16	2021 الربع الأول	6,051,404	Non-Saudi	Male
17	2021 الربع الثاني	5,869,394	Non-Saudi	Male
18	2021 الربع الثالث	5,762,323	Non-Saudi	Male
19	2021 الربع الرابع	6,010,505	Non-Saudi	Male
20	2022 الربع الأول	6,424,480	Non-Saudi	Male
21	2022 الربع الثاني	6,787,008	Non-Saudi	Male
22	2022 الربع الثالث	6,955,296	Non-Saudi	Male
23	2022 الربع الرابع	7,019,759	Non-Saudi	Male
24	2021 الربع الأول	250,388	Non-Saudi	Female
25	2021 الربع الثاني	255,438	Non-Saudi	Female
26	2021 الربع الثالث	260,754	Non-Saudi	Female
27	2021 الربع الرابع	279,991	Non-Saudi	Female

Then, split the values in the "Citizenship and Gender" column into two separate columns and drops the original column.

[11]:
```
gender_dim = pd.read_csv('Emp_Gender_Dim.csv')

gender_dim
```

[11]:

	Gender_ID	Gender
0	1	Male
1	2	Female

[12]:
```
fact_table = pd.merge(fact_table, gender_dim, on='Gender', how='left')

fact_table
```

[12]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender	Gender_ID
0	الربع الأول 2021	1,365,654	Saudi	Male	1
1	الربع الثاني 2021	1,385,268	Saudi	Male	1
2	الربع الثالث 2021	1,416,888	Saudi	Male	1
3	الربع الرابع 2021	1,469,850	Saudi	Male	1
4	الربع الأول 2022	1,531,720	Saudi	Male	1
5	الربع الثاني 2022	1,563,771	Saudi	Male	1
6	الربع الثالث 2022	1,582,946	Saudi	Male	1
7	الربع الرابع 2022	1,611,085	Saudi	Male	1
8	الربع الأول 2021	723,789	Saudi	Female	2
9	الربع الثاني 2021	680,070	Saudi	Female	2
10	الربع الثالث 2021	718,420	Saudi	Female	2
11	الربع الرابع 2021	770,962	Saudi	Female	2
12	الربع الأول 2022	841,770	Saudi	Female	2
13	الربع الثاني 2022	879,182	Saudi	Female	2
14	الربع الثالث 2022	926,180	Saudi	Female	2
15	الربع الرابع 2022	970,330	Saudi	Female	2
16	الربع الأول 2021	6,051,404	Non-Saudi	Male	1
17	الربع الثاني 2021	5,869,394	Non-Saudi	Male	1
18	الربع الثالث 2021	5,762,323	Non-Saudi	Male	1
19	الربع الرابع 2021	6,010,505	Non-Saudi	Male	1
20	الربع الأول 2022	6,424,480	Non-Saudi	Male	1
21	الربع الثاني 2022	6,787,008	Non-Saudi	Male	1
22	الربع الثالث 2022	6,955,296	Non-Saudi	Male	1
23	الربع الرابع 2022	7,019,759	Non-Saudi	Male	1

Merging the Employment Fact table with the Gender Dimension.

18	19	2022	Q3	July
19	20	2022	Q3	August
20	21	2022	Q3	September
21	22	2022	Q4	October
22	23	2022	Q4	November
23	24	2022	Q4	December

[16]:
```
date_dim.drop("Month", axis=1, inplace=True)

date_dim
```

[16]:

	Date_ID	Year	Quarter
0	1	2021	Q1
1	2	2021	Q1
2	3	2021	Q1
3	4	2021	Q2
4	5	2021	Q2
5	6	2021	Q2
6	7	2021	Q3
7	8	2021	Q3
8	9	2021	Q3
9	10	2021	Q4
10	11	2021	Q4
11	12	2021	Q4
12	13	2022	Q1
13	14	2022	Q1
14	15	2022	Q1
15	16	2022	Q2
16	17	2022	Q2
17	18	2022	Q2
18	19	2022	Q3
19	20	2022	Q3
20	21	2022	Q3
21	22	2022	Q4
22	23	2022	Q4
23	24	2022	Q4

Dropping the "Month" column from the DataFrame ‘date_dim’ because the Employment dataset is indexed by quarters, not months like the Tourism dataset.

```
[17]: date_dim.drop_duplicates(subset=["Year", "Quarter"], inplace=True)

      date_dim
```

[17]:

	Date_ID	Year	Quarter
0	1	2021	Q1
3	4	2021	Q2
6	7	2021	Q3
9	10	2021	Q4
12	13	2022	Q1
15	16	2022	Q2
18	19	2022	Q3
21	22	2022	Q4

Remove duplicate rows that appeared after dropping the "Month" column from the DataFrame.

```
[19]: fact_table['Year'] = fact_table['Year - Quarter'].str.extract(r'(\d{4})')

      fact_table
```

[19]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender	Gender_ID	Citizenship_ID	Year
0	2021 الربع الأول	1,365,654	Saudi	Male	1	1	2021
1	2021 الربع الثاني	1,385,268	Saudi	Male	1	1	2021

```
[20]: fact_table['Quarter'] = fact_table['Year - Quarter'].str.extract(r'الربع (\S+)')

      fact_table
```

[20]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender	Gender_ID	Citizenship_ID	Year	Quarter
0	2021 الربع الأول	1,365,654	Saudi	Male	1	1	2021	الأول
1	2021 الربع الثاني	1,385,268	Saudi	Male	1	1	2021	الثاني
2	2021 الربع الثالث	1,416,888	Saudi	Male	1	1	2021	الثالث

```
[21]: quarter_mapping = {
          'الأول': 'Q1',
          'الثاني': 'Q2',
          'الثالث': 'Q3',
          'الرابع': 'Q4'
      }
      fact_table.loc[:, 'Quarter'] = fact_table['Quarter'].map(quarter_mapping)

      fact_table
```

[21]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender	Gender_ID	Citizenship_ID	Year	Quarter
0	2021 الربع الأول	1,365,654	Saudi	Male	1	1	2021	Q1
1	2021 الربع الثاني	1,385,268	Saudi	Male	1	1	2021	Q2
2	2021 الربع الثالث	1,416,888	Saudi	Male	1	1	2021	Q3

```
[24]: fact_table = fact_table.merge(date_dim, on=['Year', 'Quarter'], how='left')

      fact_table
```

[24]:

	Year - Quarter	Number_of_Employment	Citizenship	Gender	Gender_ID	Citizenship_ID	Year	Quarter	Date_ID
0	2021 الربع الأول	1,365,654	Saudi	Male	1	1	2021	Q1	1
1	2021 الربع الثاني	1,385,268	Saudi	Male	1	1	2021	Q2	4
2	2021 الربع الثالث	1,416,888	Saudi	Male	1	1	2021	Q3	7
3	2021 الربع الرابع	1,469,850	Saudi	Male	1	1	2021	Q4	10
4	2022 الربع الأول	1,531,720	Saudi	Male	1	1	2022	Q1	13

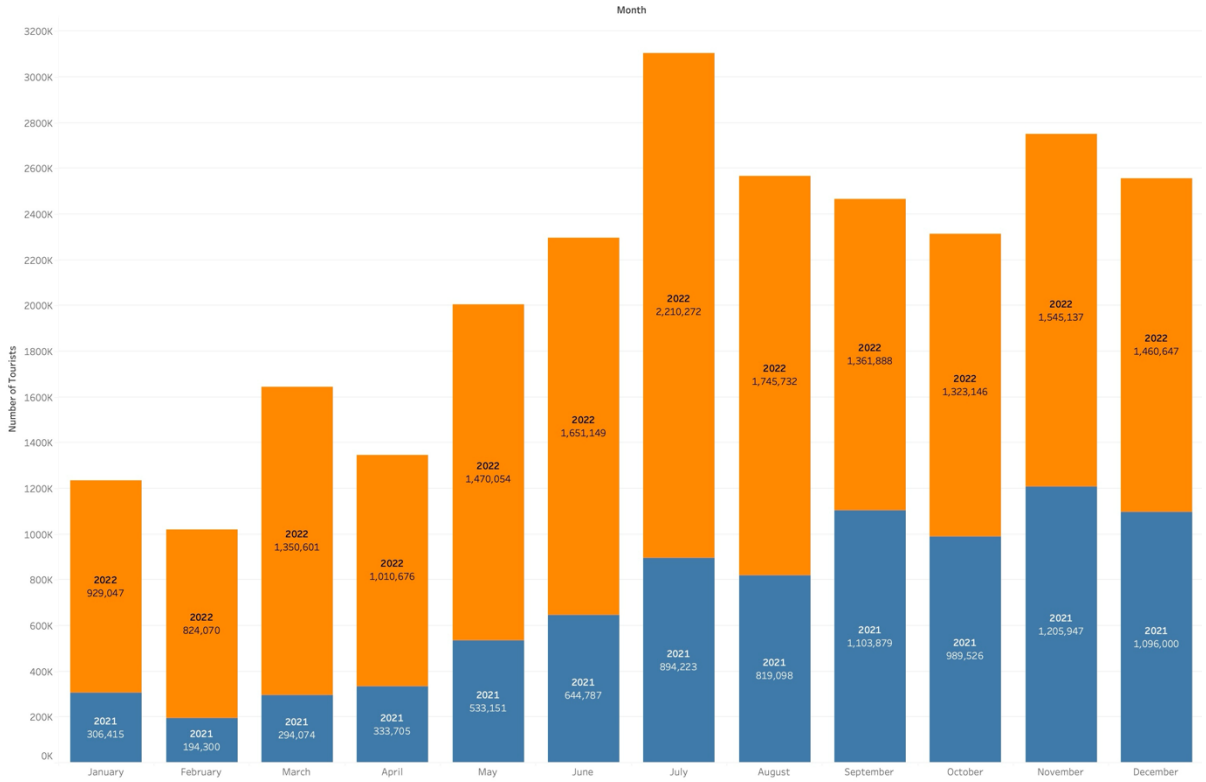
Modifying Year and Quarter columns to match Date Dimension format.

Phase#5: Data Visualizations

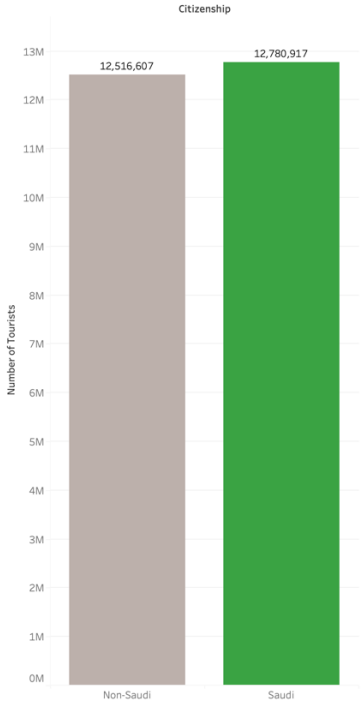
Here we'll show the visualization for each KPI, we used tableau to create them.

For the Tourism Dataset KPIs:

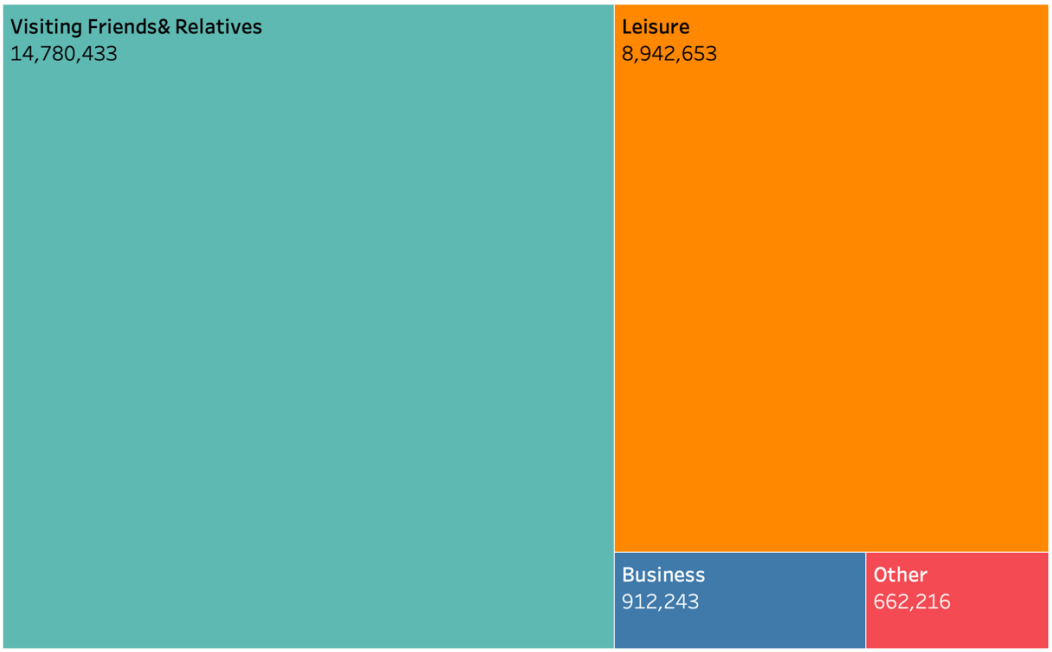
1. Number of Tourists per Month



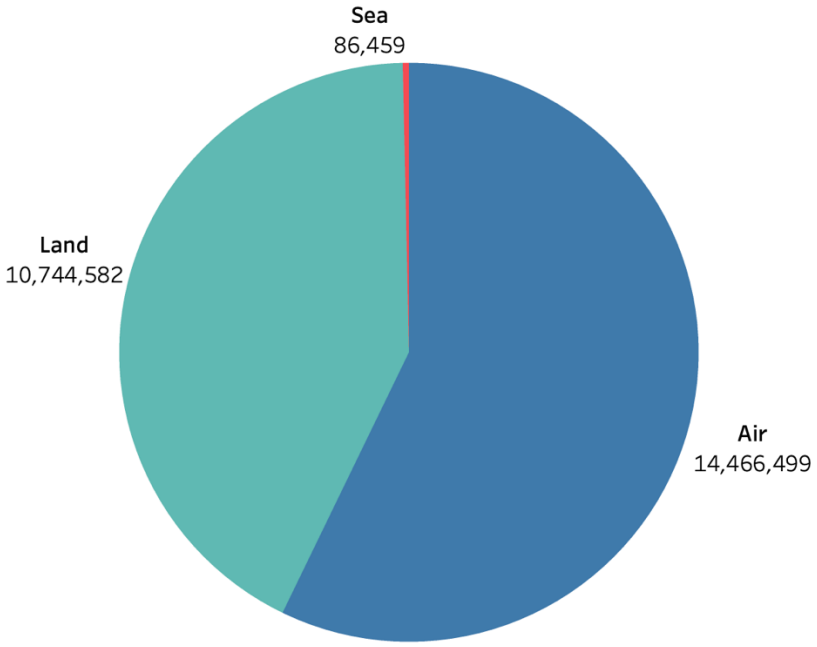
2. Number of Tourists by Citizenship



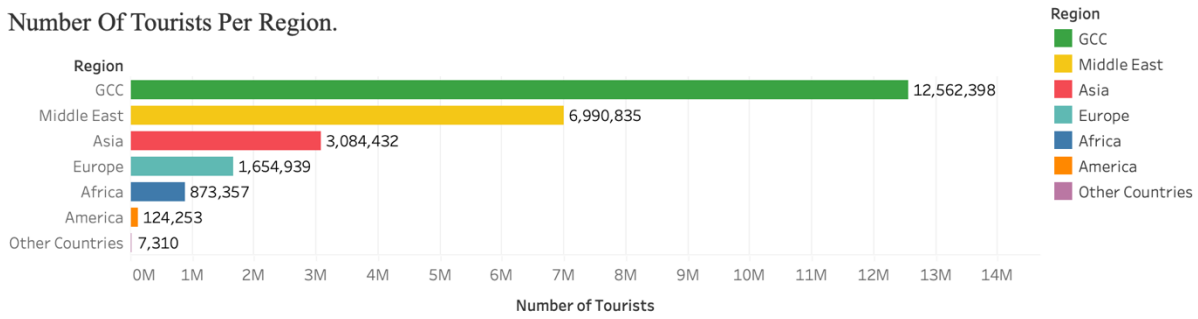
3. Number of Tourists by Purpose of Trips



4. Number of Tourists by Mode of Departure

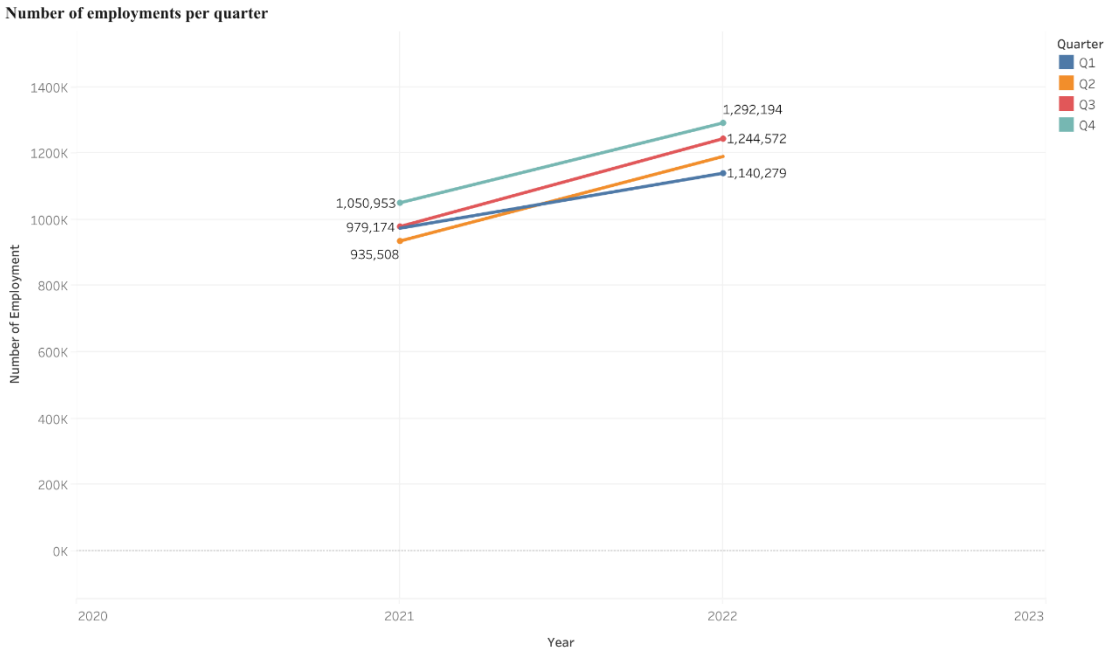


5. Number of Tourists per Region

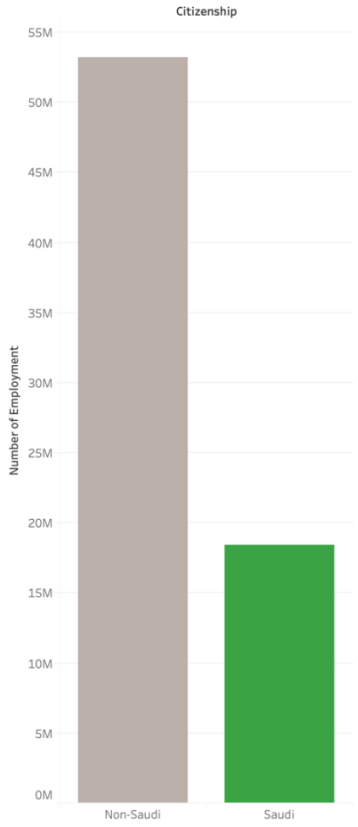


For the Employment Dataset KPIs:

1. Number of Employments per Quarter

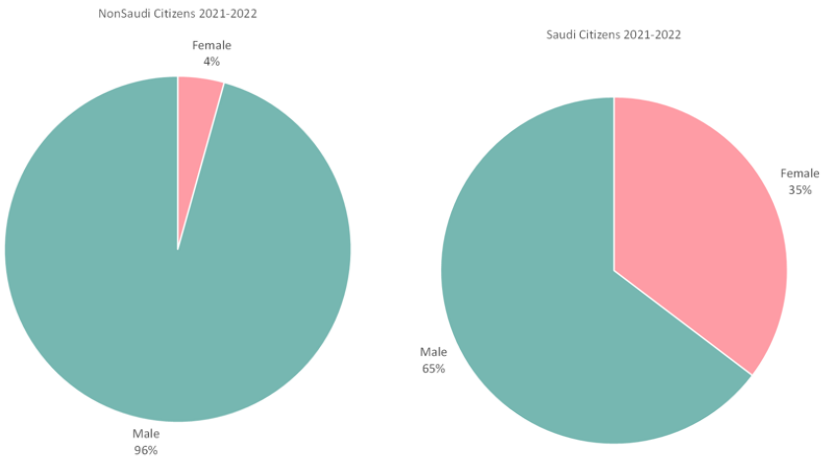


2. Number of Employments by Citizenship

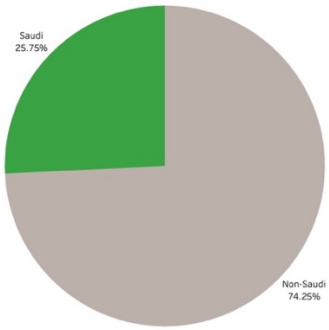


3. Number of Employments per Gender

4. Ratio of Gender by Citizenship

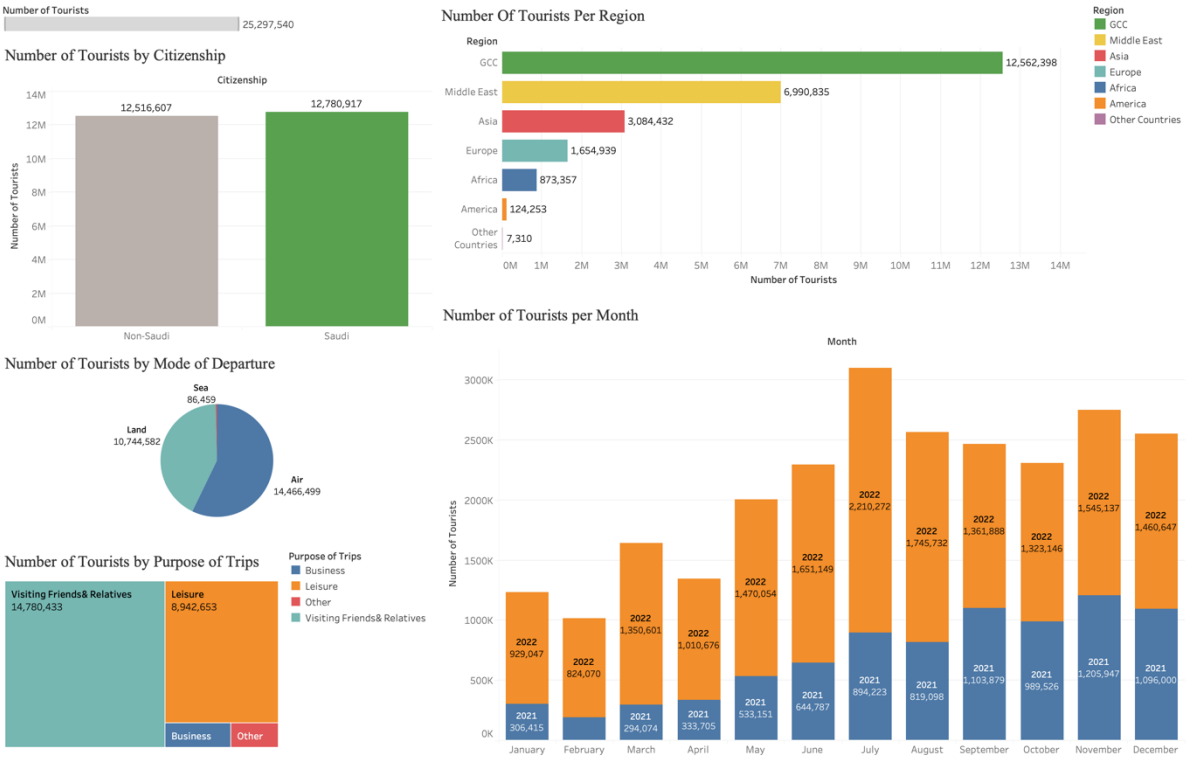


5. Ratio of Saudi to Non-Saudi Employment

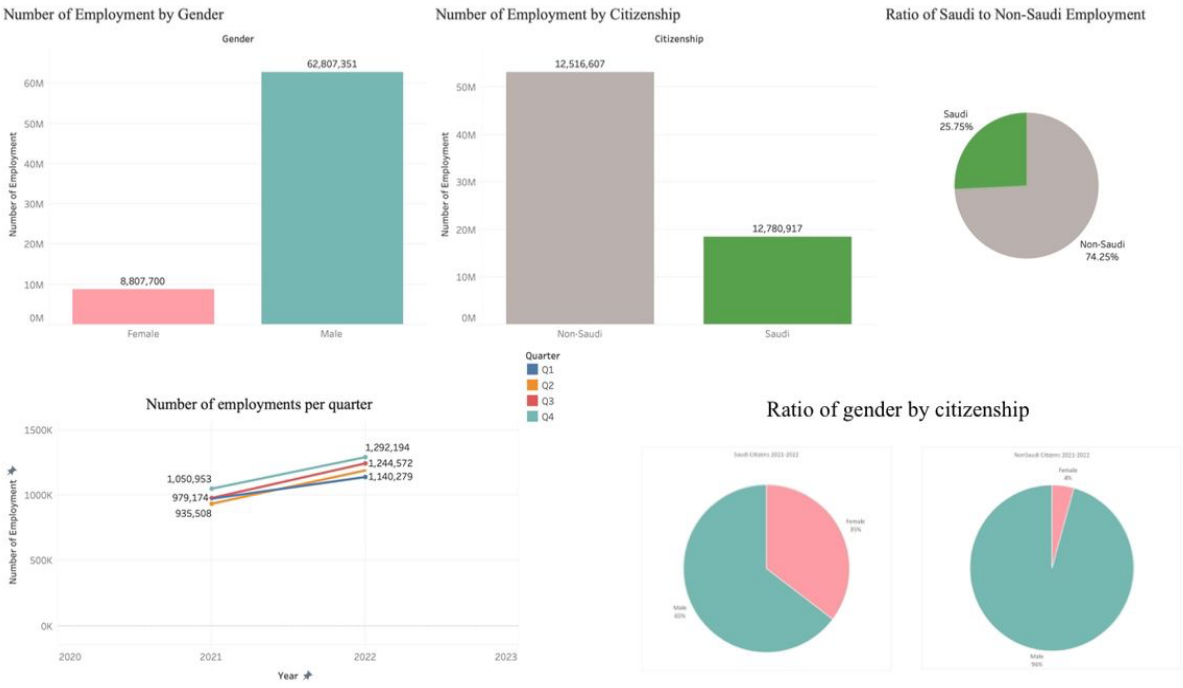


Dashboard for Each Dataset:

- Tourism Dashboard



- Employment Dashboard



Conclusion:

The analysis of the dashboards reveals no correlation between employment and tourism. Our data shows an increasing trend in the number of outbound tourists over time, particularly peaking in July. Although there are monthly fluctuations due to various unknown factors, the overall yearly trend in tourism is upward, with most trips being made to visit friends and relatives. Additionally, there is no significant difference in the number of Saudi versus non-Saudi tourists. Most tourists depart by air, and the GCC region receives more tourists compared to other regions.

The employment data indicates an increase in the number of employees during the fourth quarter, which could be attributed to seasonal and economic factors. The data also shows a higher percentage of male employees compared to female employees. Furthermore, a significant portion of the workforce is composed of non-Saudis, reflecting Saudi Arabia's reliance on foreign labor to meet certain skill requirements.

Work Distribution:

Name	ID	Percentage	Work
Rand Aljathlani		100%	- Problem Formulation - Identify Dataset - KPI's - Data Modeling - ETL Process - Visualization - Conclusion
Fulwah Aldhuwayhi		100%	- Problem Formulation - Identify Dataset - KPI's - Data Modeling - ETL Process - Visualization - Conclusion
Alhanouf Alhayan		100%	- Problem Formulation - Identify Dataset - KPI's - Data Modeling - ETL Process - Visualization - Conclusion
Leena Alateeq		100%	- Problem Formulation - Identify Dataset - KPI's - Data Modeling - ETL Process - Visualization - Conclusion
Sadeem Almushham		100%	- Problem Formulation - Identify Dataset - KPI's - Data Modeling - ETL Process - Visualization - Conclusion