CBA | College Of Business Administration كلية إدارة الأعمال

جامعة الإمام عبدالرحمن بن فيصل
IMAM ABDULRAHMAN BIN FAISAL UNIVERSITY

Course title: Business intelligence

Instructor: Ms. Moneerah Albuolyan

Final Project

Section:3208

Group 8

ID	Student name
2210002496	Torkiah Jamal Alharthi
2210003293	Rafaa Yagoob Alghamdi
2210002794	Aliah Hammad Alshammari
2210006230	Ghada Turki Almutairi

Contribution Table

Student's name		Contributed part
1.	Torkiah Jamal Alharthi	Identify the business problem definition and ensure data preparation
2.	Rafaa Yagoob Alghamdi	Analyzing the population data set through orange and extracting the population prediction
3.	Aliah Hammad Alshammari	Interpret and comment on the results
4.	Ghada Turki Almutairi	Descriptive analysis of historical data and description of the technique used.

- **Business problem definition**

The business problem is concerned with population prediction. For our dataset, we want to predict the population of Saudi and non-Saudi citizens in Saudi Arabia by region via linear regression and ensure whether our predictions are accurate or not. The analysis and prediction of the population are based on different factors such as gender, nationality, region, and years which are related to 2016 until 2022 past data.

- **Data preparation**

In our dataset, there are no missing values, where data is free from noise and inconsistencies. So, the data is in a suitable structure for analysis. Moreover, we have one target (the population) and 4 features (gender, nationality, region, and years)

- **Descriptive analysis of historical data**

This analysis explores an extensive dataset that precisely records Saudi Arabia's population evolution between 2010 and 2022. The dataset has been constructed up of 676 rows, each of which corresponds to a distinct record that records demographic data for Saudi Arabia. Multiple attributes or columns within these rows carefully detail important demographic markers. Important characteristics are nationality, which separates Saudi citizens from non-Saudi individuals, and gender, which shows the distribution of male and female citizens, and geographic distribution throughout all thirteen regions—Al Bahah, Al Jawf, Al Madinah Al Munawwarah, Al Qaseem, Ar Riyadh, Aseer, Eastern Region, Hail, Jazan, Makkah Al Mukarramah, Najran, Northern Borders, and Tabuk—are just a few of the nuanced perspectives on the dynamics of population that these observations offer. Each attribute functions as a crucial dimension for assessing and comprehending how complicated relationships exist between demographic elements in Saudi Arabia, with a structure consisting of 676 rows and 5 columns. This allows for a comprehensive analysis of population patterns throughout the designated period.

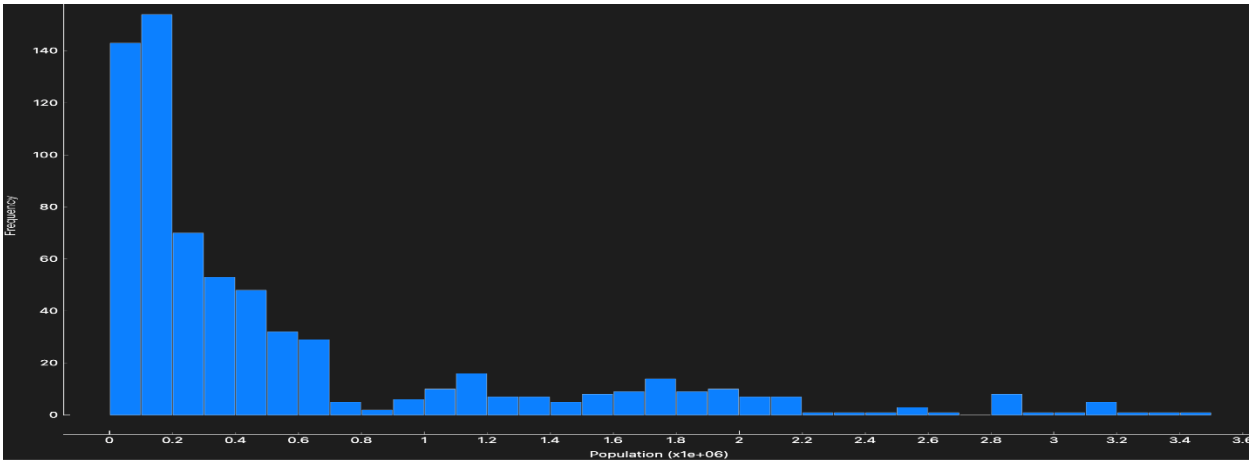

Figure 1.0

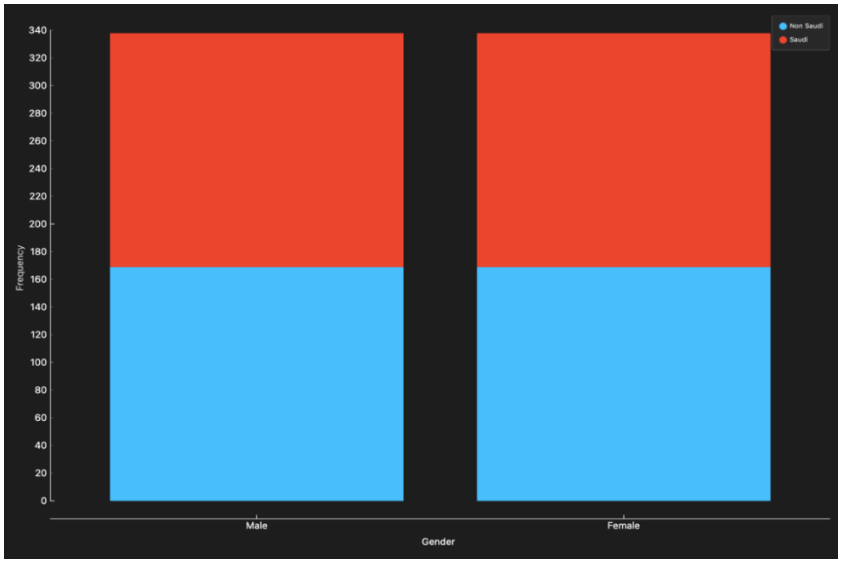

Figure 1.1

With important statistical metrics providing insightful information, the graph in figure (1.0) vividly depicts changes in population throughout time. The mean population across all year's value of 558,716.90 highlights the extensive character of the dataset, highlighting its breadth. Notably, certain concentrations across this demographic range are highlighted by the mode at 16,230 and the median at 249,559. The population's range, which ranges from 14,304 minimum to 3,406,281 maximum individuals, demonstrates the dataset's breadth. The computed dispersion of 1.26 provides more insight into the variability of the dataset and suggests that there may be underlying demographic complexities throughout time,

Furthermore, the population's positively skewed distribution emphasizes the dataset's comprehensive portrayal of demographic patterns, which often include a variety of factors including migration, birth rates, death rates, and more.

Referring to figure (1.1) the population's gender and nationality distributions across the years of 2010-2022 are extremely balanced. There are exactly equal numbers of Saudi and non-Saudi males, Saudi females, and non-Saudi females—169 total—representing each gender. It's interesting to note that each gender contributes equally, making up 50% of the population. Additionally, the data show that the percentage of Saudi and non-Saudi citizens are balanced, with each group making up exactly 50%. A significant demographic symmetry in the dataset is highlighted by the balance in the distribution of genders and nationalities, which reflects a carefully managed balance in the population composition.

- **Description of the technique used**

Data instances: 19
Features: 1
Meta attributes: 1

	name	coef
1	intercept	-1.25743e+19
19	Year	21504
4	Nationality=Non Saudi	3.93422e+17
5	Nationality=Saudi	3.93422e+17
6	Region=Al Bahah	3.37971e+18
17	Region=Northern Borders	3.37971e+18
16	Region=Najran	3.37971e+18
7	Region=Al Jawf	3.37971e+18
13	Region=Hail	3.37971e+18
18	Region=Tabuk	3.37971e+18
9	Region=Al Qaseem	3.37971e+18
14	Region=Jazan	3.37971e+18
11	Region=Aseer	3.37971e+18
8	Region=Al Madinah Al Munawwarah	3.37971e+18
12	Region=Eastern Region	3.37971e+18
10	Region=Ar Riyadh	3.37971e+18
15	Region=Makkah Al Mukarramah	3.37971e+18
2	Gender=Female	8.80119e+18
		8.80119e+18

Figure 2.0

Data instances: 338
Features: 4
Meta attributes: 2
Target: Numeric variable 'Population'

	Population	LR	LR (error)	Gender	Nationality	Region	Year
1	85946	194560	108614	Male	Non Saudi	Al Bahah	2016
2	134730	253952	119222	Male	Non Saudi	Al Jawf	2016
3	666678	575488	91190	Male	Non Saudi	Al Madinah Al Munawwarah	2016
4	394872	411648	16776	Male	Non Saudi	Al Qaseem	2016
5	3406281	1.86982e+06	1.53646e+06	Male	Non Saudi	Ar Riyadh	2016
6	504117	536576	32459	Male	Non Saudi	Aseer	2016
7	1861202	1.14893e+06	712274	Male	Non Saudi	Eastern Region	2016
8	178799	278528	99729	Male	Non Saudi	Hail	2016
9	338452	419840	81388	Male	Non Saudi	Jazan	2016
10	3279865	1.87802e+06	1.40185e+06	Male	Non Saudi	Makkah Al Mukarramah	2016
		249856	93856	Male	Non Saudi	Najran	2016

Figure 2.1

We chose to use the prediction method—more particularly linear regression—because we needed to predict future population changes of Saudi and non-Saudi citizens in Saudi Arabia using past data. Our goal was to record any possible shifts or patterns in population patterns throughout time, therefore we divided the dataset into two separate periods: 2010-2016 and 2016–2022.
By using of the powerful statistical technique known as linear regression, we may accurately predict the connection between independent factors (for instance, time) and dependent variables (population size). Through the application of this method, we can conclude future population estimates by extrapolating patterns from historical data.
The results of a linear regression (LR) model are shown in the first data table (Figure 2.0), which provides information about the coefficients associated with different predictor variables. It consists of 19 instances, each including a distinct region, year, gender, and nationality, together with a corresponding coefficient number. These coefficients represent the degree to which each predictor variable affects population estimate.
There is a noticeable negative bias in the interception coefficient, which suggests that there are possible problems with the baseline estimates of the model. On the other hand, coefficients related to certain variables, such as year and gender, show significant positive values, indicating their major effect on population estimates. Furthermore, coefficients associated with nationality and region exhibit a great deal of variability, suggesting substantial demographic disparities among Saudi and non-Saudi populations, and across different regions.
In Figure (2.1), This table provides a brief glimpse of the comparison between the anticipated and actual population numbers. With 338 instances, it illustrates particular demographic circumstances based on region, year, gender, and nationality. By comparing the expected and actual population data, it is possible to assess how accurate and dependable the model is in predicting demographic changes over the given period. It sheds light on the linear regression model's prediction accuracy. When it comes to some years, like 2016 for example, the model quite much matches real population estimates. However, differences between estimated and actual values are visible in some circumstances, one prominent example of where differences are evident is concerning the 2018 demographic forecast, suggesting that the model has limits or opportunities for development. These differences highlight how difficult it is to predict demographic changes and how crucial it is to improve prediction models to appropriately reflect changing population patterns.

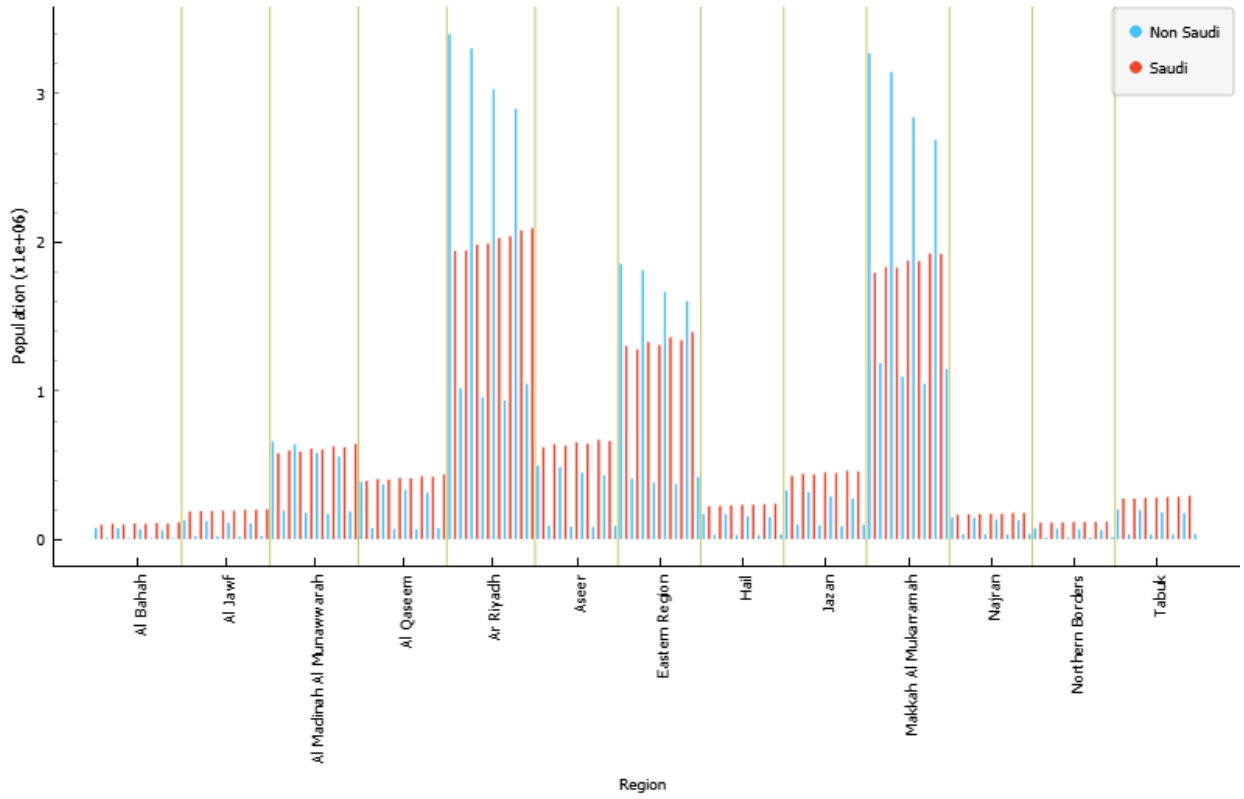

Figure 3.1

- **Results interpretation**

The bar chart in figure 3.1 is a visual representation of the population growth prediction by region for the years 2016 – 2022. We can conclude from it the regions with the highest expected population increase and the lowest in addition to the nationalities of citizens of said region. Riyadh and Mecca have the highest expectancy of population increase mostly by non-Saudi citizens. The regions with the least expected population flourish are El Baha and the Northern borders with mostly Saudi citizens. The difference in ratio between Saudi and non-Saudi citizens may be due to different reasons for example religious reasons. Mecca, is the global destination for Muslims that come for Hajj and Umrah (GOV, 2024). As for cities like Riyadh and the Eastern region, job opportunities play a major role in the population boost. Governmental decisions are similar to the new Saudi contract law that requires firms to move their regional headquarters to Saudi Arabia (SaudiGazette, 2023), in addition to having companies that are industry giants like Saudi Aramco and Sabic help produce major opportunities and therefore

increase population. Other regions like El Medina manora and Asser also expect an approximately equal population increase of both Saudi and non-Saudi citizens. Regions like Al Jawf, Al Qaseem, Hail, Jazan, Najran, and Tabuk expect a moderate increase in population of both Saudi and non-Saudi citizens overall these regions are considered rural regions of Saudi Arabia most of its citizens are people whose tribes owned the place pre founding of Saudi Arabia or are retired and are looking for a slower paced city than Riyadh and the Eastern Region. Overall, the population of Saudi Arabia is expected to increase in the upcoming years of 2016 - 2022.

References:

SaudiGazette (2023) *200+ firms move HQS to Riyadh ahead of Saudi Govt Contract Policy*, *Saudigazette*. Available at: https://www.saudigazette.com.sa/article/639215 (Accessed: 06 May 2024).

GOV (2024) *Hajj statistics publication*, *General Authority for Statistics*. Available at: https://www.stats.gov.sa/en/28 (Accessed: 06 May 2024).