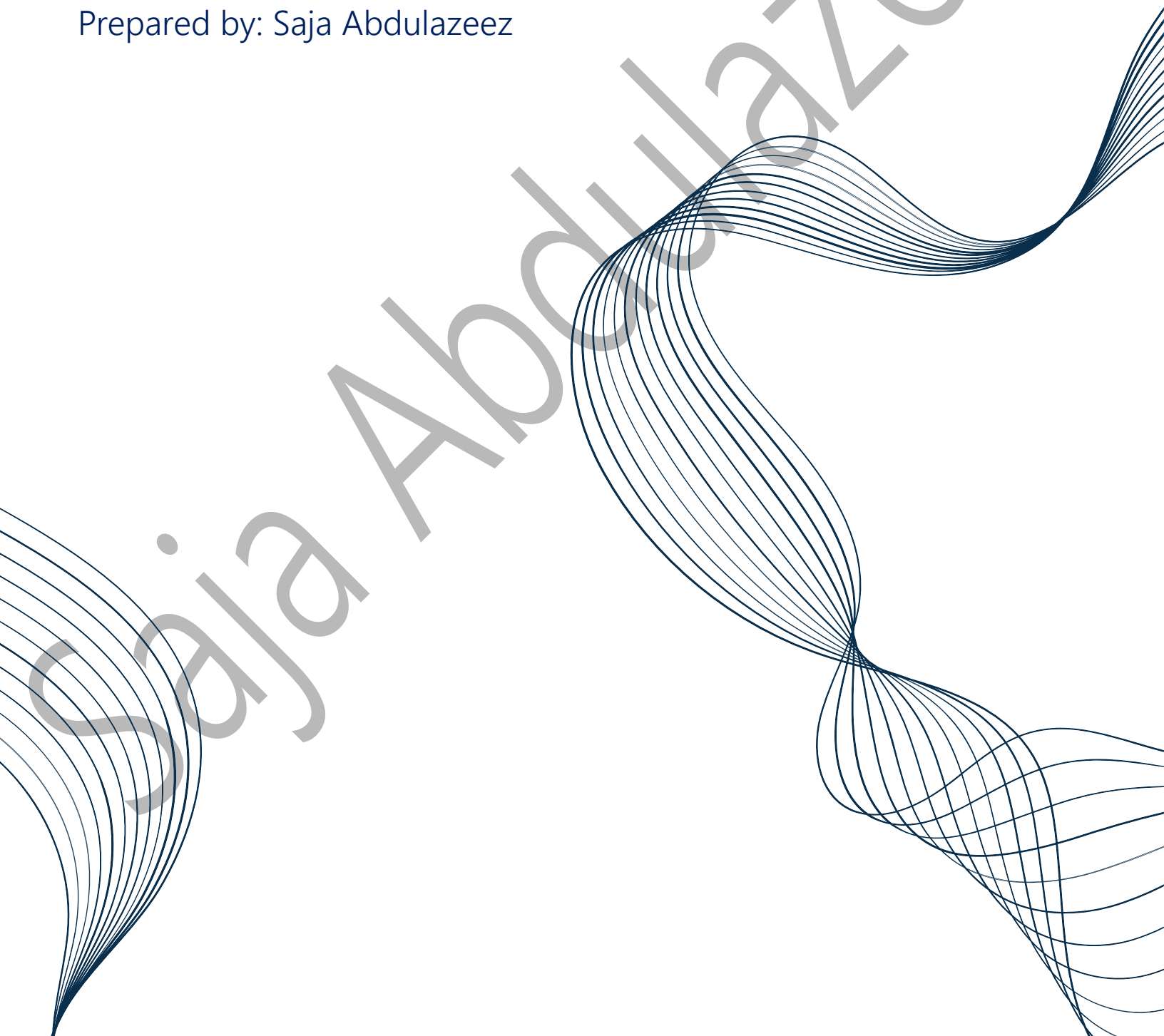
Big Data Analytics and Visualization

Prepared by: Saja Abdulazeez



Agriculture in Jordan

From the Jordan valley to Irbid and Ghor Al-Safi, agriculture holds a great significance in Jordan, harvesting olives and tending to livestock are traditions deep rooted within our culture. This country's unique geographical location and climatic conditions is what helped to shape the agricultural landscape we know today, utilizing open lands, rain-fed, and irrigation farming methods. However, due to climate change, the weather has been unpredictable, and farmers are finding it difficult to improve their land use and are often disappointed with crop production, this paper covers weather-based crop prediction for the year 2022, in order to avoid losses and make better decisions on which crops to grow.

The Agricultural Market

Agriculture plays a fundamental role in providing food security and economic stability in the kingdom and outside of it. Since 2010 we notice a surge in exports of fruits and vegetables to the European market, reaching a total of €651 million in 2015 with tomatoes accounting for 65% of all vegetable exports, additionally in that year, agricultural exports have shifted towards the gulf market with UAE being in the lead[1]

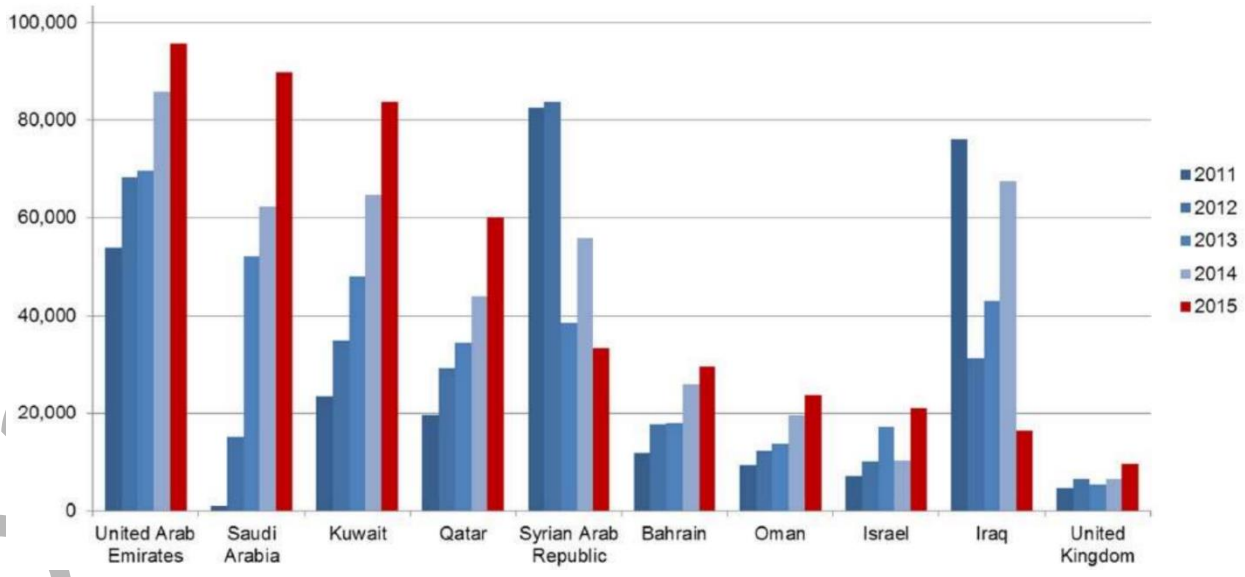


Figure 1 Jordanian vegetable exports (HS07): main markets (Leeters & Rikken, 2016)

Consider the following table for the main agricultural exports from Jordan in 2017[2]:

Commodity	,000 US$	MT unless otherwise noted
Tomatoes	223,054	282,271
Live sheep (number)	161,827	497,091 head
Peppers	56,068	47,970
Livestock forage	36,395	30,857
Cheese	28,034	6,436
Squash	23,372	27,693
Sweet melon	15,034	35,417
Cucumbers	11,545	19,024
Watermelons	10,424	19,095
Poultry meat	9,998	5,034
Cauliflower	9,717	14,414
Eggs (number)	7,187	34,055,400 eggs

Main Challenges

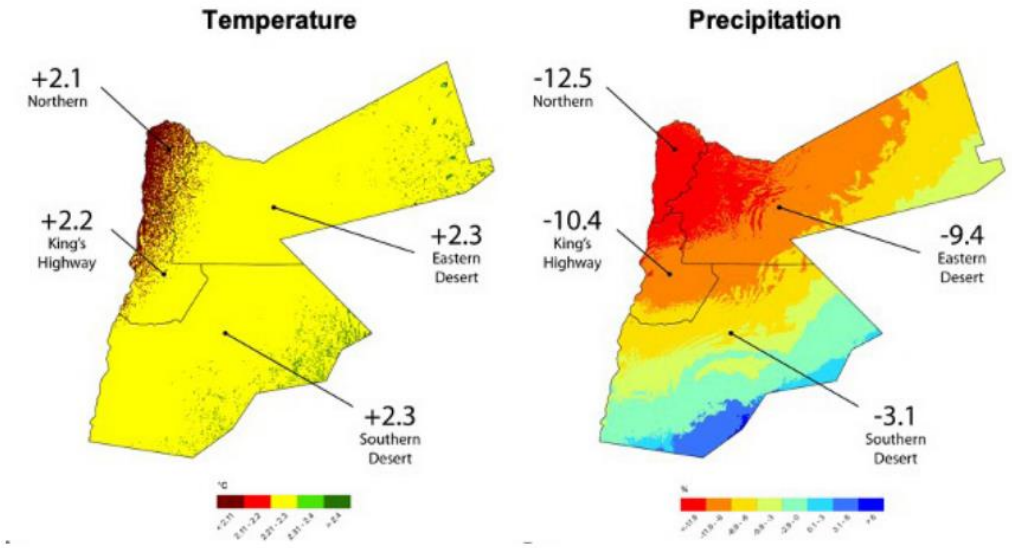
It's clear that Jordan plays a major role in agricultural exports, however, reaching those numbers is no easy task, agriculture in Jordan has faced (and currently is facing) some major challenges including but not limited to:

- The unpredictable weather due to climate change has negatively affected the agricultural sector, causing a drop in some crop yields and over-production in others. [3]
- Majority of crop production systems in Jordan are open field and depend on rainfall and stores of rainwater, which means crop production can be heavily influenced by the weather. 31% of Jordan Valley's vegetable production systems are open field[1] compared to only 11% which are in controlled greenhouse environments
- On august 31st 2022, the kingdom witnessed a 10-day heat wave rising +15° than normal, decreasing tomatoes in the central market by 50% in that year. [4]

- This increase in temperature has drastically affected precipitation levels causing drastic droughts across the kingdom and unfortunately this is expected to steadily increase by 2030.

Figure ES.2 Projected changes in annual mean temperature and total annual precipitation in Jordan

Projected changes by 2030 for Representative Concentration Pathway 8.5 (high emissions



This drastic shift in weather is out of our hands we need to find a way to adapt to this change, with that being said, can we find a way to predict the yield of crops based on historical weather data, to aid the decision-making process for agricultural exports for the year 2022?

Strategy

Our main business goal is to stay ahead of the market, with today's technologies and abundance of data we can indeed gain confidence in making decisions regarding what crops to focus on producing for the upcoming years (2022). This can be achieved by using big data tools to analyze historical multivariate timeseries weather data collected by Arabia Weather and crop yields across the years from the Department of Statistics data bank to discover a relationship between weather and crop yield. With this relationship we are able to predict the yield of crops in the upcoming years with machine learning in PySpark.

About the datasets

As mentioned briefly above, we have been given 3 data sets, 2 historical weather datasets from 2017-2023 for regions Irbid and Ghor El Safi, and a crop production data set for the respective regions from the years 2017-2021

- Crop production: the data bank contains data of 27 crops and their productions from the years 1994-2021 from different regions in Jordan
    - I will approach this by choosing the regions Ghor El Safi and Irbid , from the years 2017-2021 (most recent)
    - For each year this data set classifies production into two seasons: Summer and Winter
    - Based on my research I chose the crops Tomatoes, Okra and Onion dry for the following reasons:
        - Tomatoes are the most produced and exported outside of Jordan
        - Okra plants are drought-tolerant vegetables[5] and thrive in high summer temperatures which in my opinion is using the increasing temperatures in our favor. It can withstand a temperature up to 41^0 Celsius but are best grown between 29°-35 °. [6]
        - Onion on the other hand are cold-season crops[7] and grows in temperatures ranging from 12°-22° degrees Celsius and onion seeds can germinate in temperatures as low as 2°. [8]
        - In EDA I also explored the crops Cucumbers and Broad Beans
        - Broad beans are considered a cool season crop meaning they thrive in late spring – autumn and are quite self-sufficient[9]. Best temperatures are 15.5-18.3 and will not grow in temps under 4.4 or above 23.8[10]
        - Cucumbers grow best in temps between 24-29 [11]

- Irbid and Ghor Safi: these two datasets contain real time historical weather information that are collected every 5-6 hours daily from the years 2017-2023, the readings include:
    - Station: this could be Irbid or Ghor El Safi
    - Date/Time (year): these are the timestamps for when the reading was taken in the form 2017-01-05 09:00:00 indicating this reading was taken January 5th, 2017, at 9AM
    - Air Dew Point: The dew point represents the level of moisture, specifically the amount of water vapor present in the air and is measured in degrees Celsius. It's based on three factors, air temperature, humidity, and atmospheric pressure. [12]
    - Air Temperature: the temperature in degrees Celsius
    - Humidity: measured in % is the percentage of humidity relative to the reading
    - Manual Present Weather: a brief description of the weather in that day
    - Cloud Type: classified in 6 main types to classify the cloud distribution they include altocumulus, cirrostratus, cirrus, cumulonimbus, cumulus, nimbostratus(only in Irbid dataset)
    - Cloud Cover (Okta): is the cloud cover in okta which is the measure of cloud cover ranging from 1-8; 1 being sparse clouds and 8 full cloud cover with no breaks[13]
    - Cloud Cover %: percentage of cloud cover
    - Wind Direction (Degrees): direction of the wind in degrees from 0-360
    - Wind Speed MPS: speed of wind in meters per second
    - Wind Type: describes the wind as either calm or normal

Exploratory Data Analysis

Now that we have a general understanding of the metadata and the datasets, in this section we'll manually explore the data and see any correlations between the features and deepen our understanding of any nulls, outliers and how we're going to handle them.

I have a separate notebook just for manual EDA in pandas however in the following steps you'll notice most of my preprocessing will be done in PySpark. I utilized the preprocessed data and saved them to csv so I can just directly read them in pandas for visuals and finding correlations. The csv's for EDA_visualizations.ipynb are "irbidForViz.csv" and "ghorForViz.csv"

The notebooks for predictions in Irbid and Ghor Safi are "ghorProduction.ipynb" and "irbidProduction.ipynb"