Saudi_Ecommerce Dataset

Supervisor:
Dr. Sameh

Student Name:
Raghad Alshalan
Wejdan Othman
Wafa Othman
Khafuq Abdullah

Content

Data Setes

This dataset foe people who are thinking of entering the e-commerce business in Saudi Arabia or just curious about the sector, this data set will come in handy. It's scrapped from Maroof website which is an initiative started by the Ministry of Commerce to help our customers to recognize official stores with legal grounds.
A high-quality eCommerce store dataset delivers extensive firmographic insights for major online retailers as well as niche players. Companies use this data to power competitor analysis and monitor competitor activities.

Introduction

E-commerce data analysis refers to the process of gathering, cleaning, transforming, and modeling data from e-commerce platforms to gain insights and make informed decisions. The goal of analyzing e-commerce data is to understand customer behavior, market trends, and business performance to drive growth and improve decision-making. E-commerce data can come from various sources, such as website analytics, customer demographics, sales data, and marketing campaigns. Then this data is processed and analyzed using various techniques, including statistical analysis, machine learning algorithms, and data visualization tools. In this project, we analyzed your collection of data from the Kaggle program using spark, and we got the best results.

Content

Each row is a Store. Columns are as follows:

- Name_ENG: Name of the store in English
- Name_AR: Name of the store in Arabic
- CR: Commercial Registration number
- Category: Store chosen category within Maroof limits
- Website: Link to the Store main selling channel
- Instagram: Store Instagram Account link
- Twitter: Store Twitter Account link
- Email: Store email address
- Phone Number: Store phone number
- Activity: Status of the business as shown in Maroof
- Rating: Store rating out of 10
- Num_Ratings: Number of ratings the store has

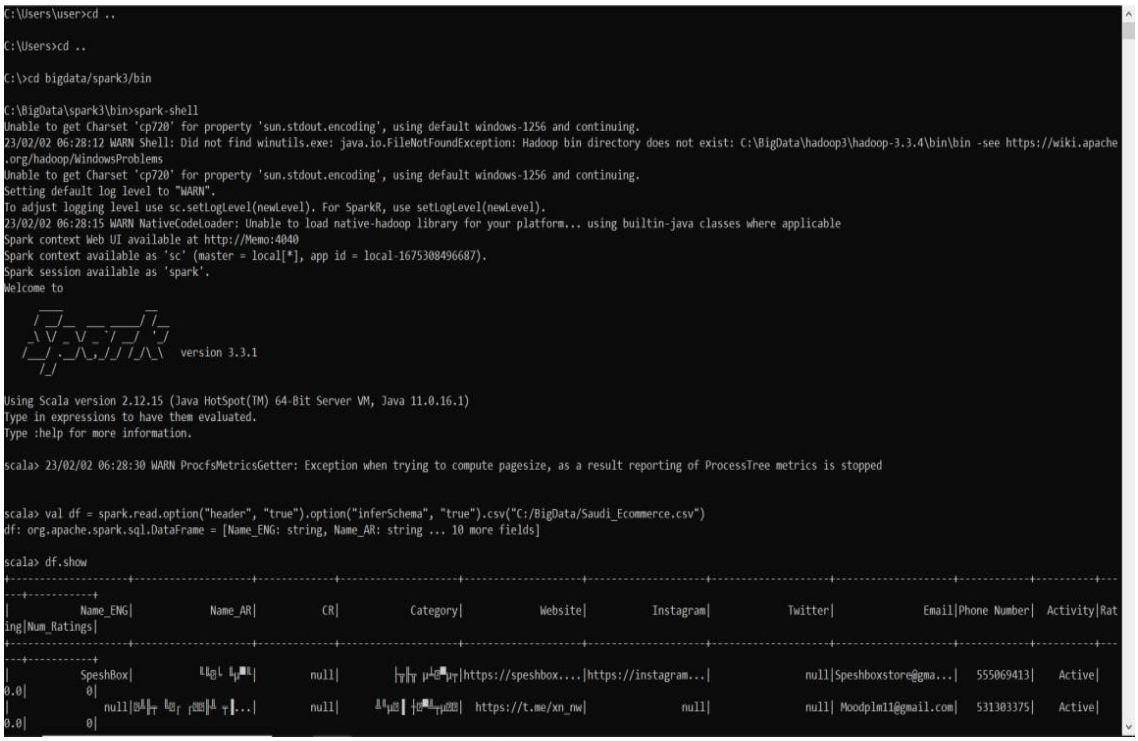
Data Type

It is Structured data. the data conforms to a data model, has a well define structure, follows a consistent order and can be easily accessed and used by a person or a computer program. it was stored in well-defined schemas (Databases). It is generally tabular with column and rows that clearly define its attributes. and can used SQL (Structured Query language) to manage this structured data stored in databases. eCommerce store data: This dataset provides details about eCommerce vendors, sellers, and platforms, creating a comprehensive database of competitors.

Benefits spark

Speed: Spark is designed to be fast, with in-memory data processing capabilities that can significantly speed up big data analysis tasks compared to traditional MapReduce.
Flexibility: Spark supports multiple data sources and data formats, including structured, semi-structured, and unstructured data.
Scalability: Spark can be easily scaled out to handle large data sets, allowing for quick and efficient processing of big data.
Ease of Use: Spark has a user-friendly API and a variety of libraries, such as Spark SQL, Spark Streaming, MLlib, and GraphX, which make it easier for data scientists and engineers to perform complex data analysis tasks.

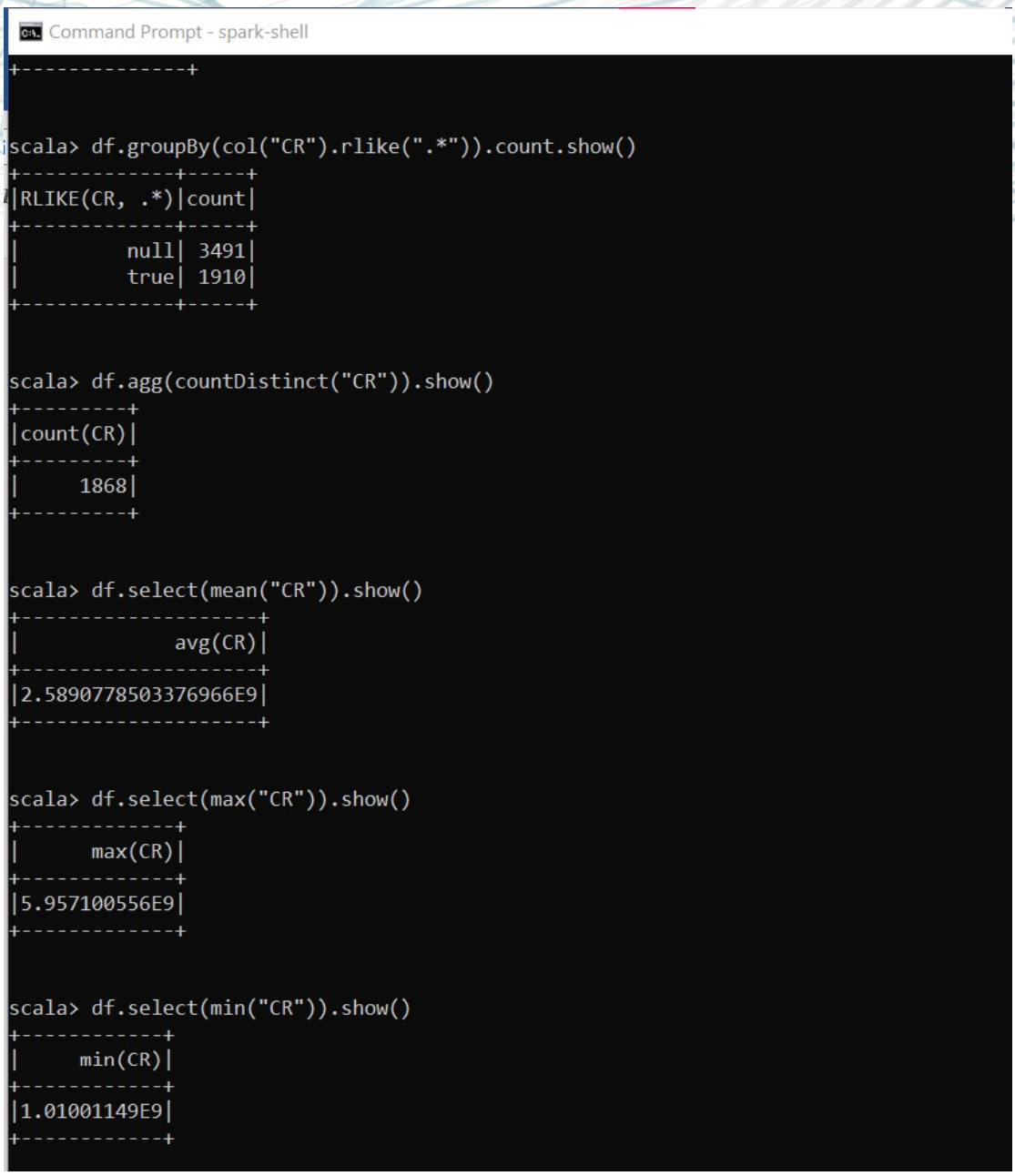
Analysis with Hadoop spark



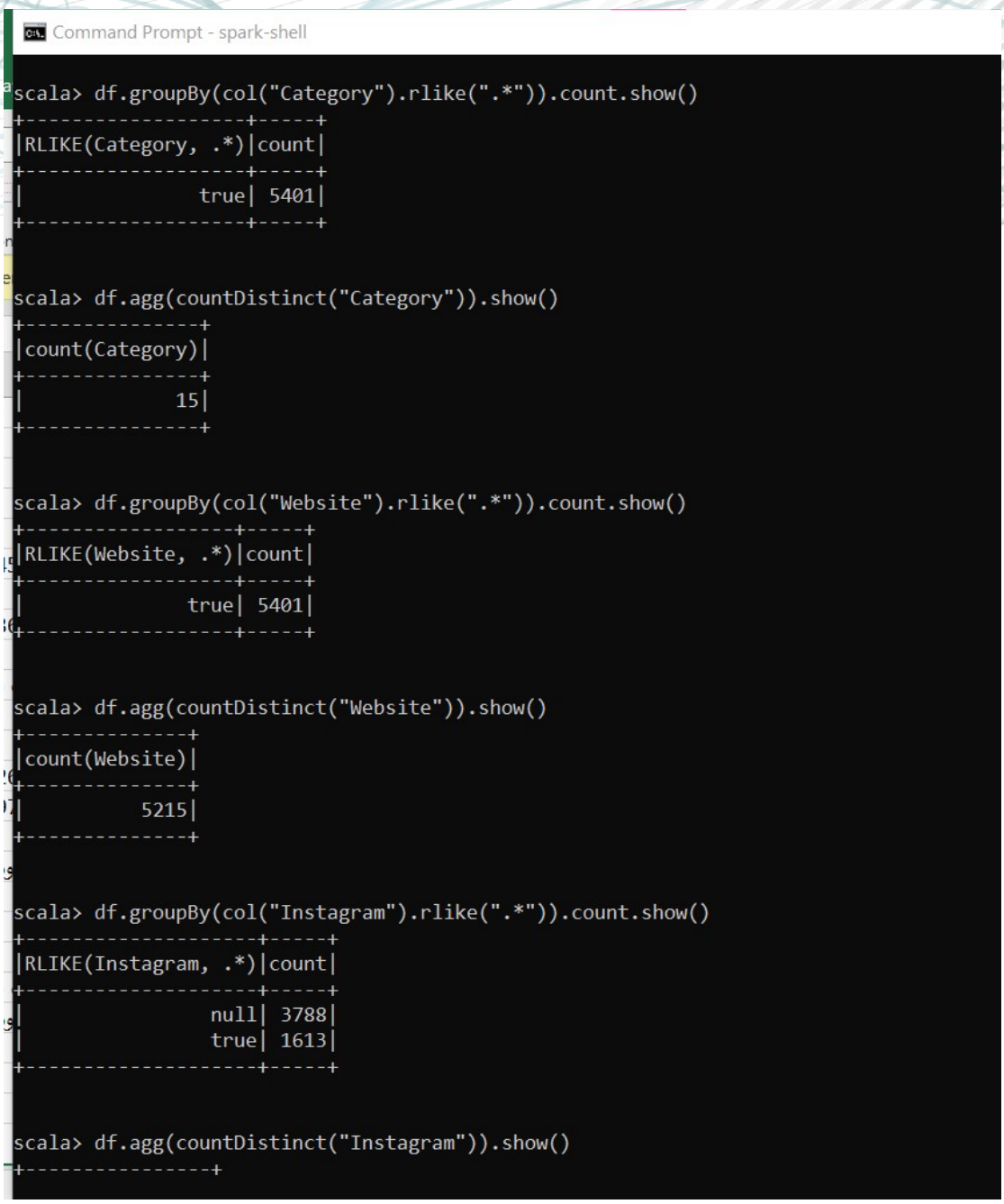
1. Getting start spark (scala).
2. Import file “Saudi_Ecommerce.csv” to spark (Data Frame) and show it.

```
scala> df.groupBy(col("Name_ENG").rlike(".*")).count.show()
+------------------+-----+
|RLIKE(Name_ENG, .*)|count|
+------------------+-----+
|              null|  391|
|              true| 5010|
+------------------+-----+


scala> df.agg(countDistinct("Name_ENG")).show()
+---------------+
|count(Name_ENG)|
+---------------+
|           5009|
+---------------+


scala> df.groupBy(col("Name_AR").rlike(".*")).count.show()
+-----------------+-----+
|RLIKE(Name_AR, .*)|count|
+-----------------+-----+
|             true| 5401|
+-----------------+-----+


scala> df.agg(countDistinct("Name_AR")).show()
+--------------+
|count(Name_AR)|
+--------------+
|          5400|
+--------------+
```

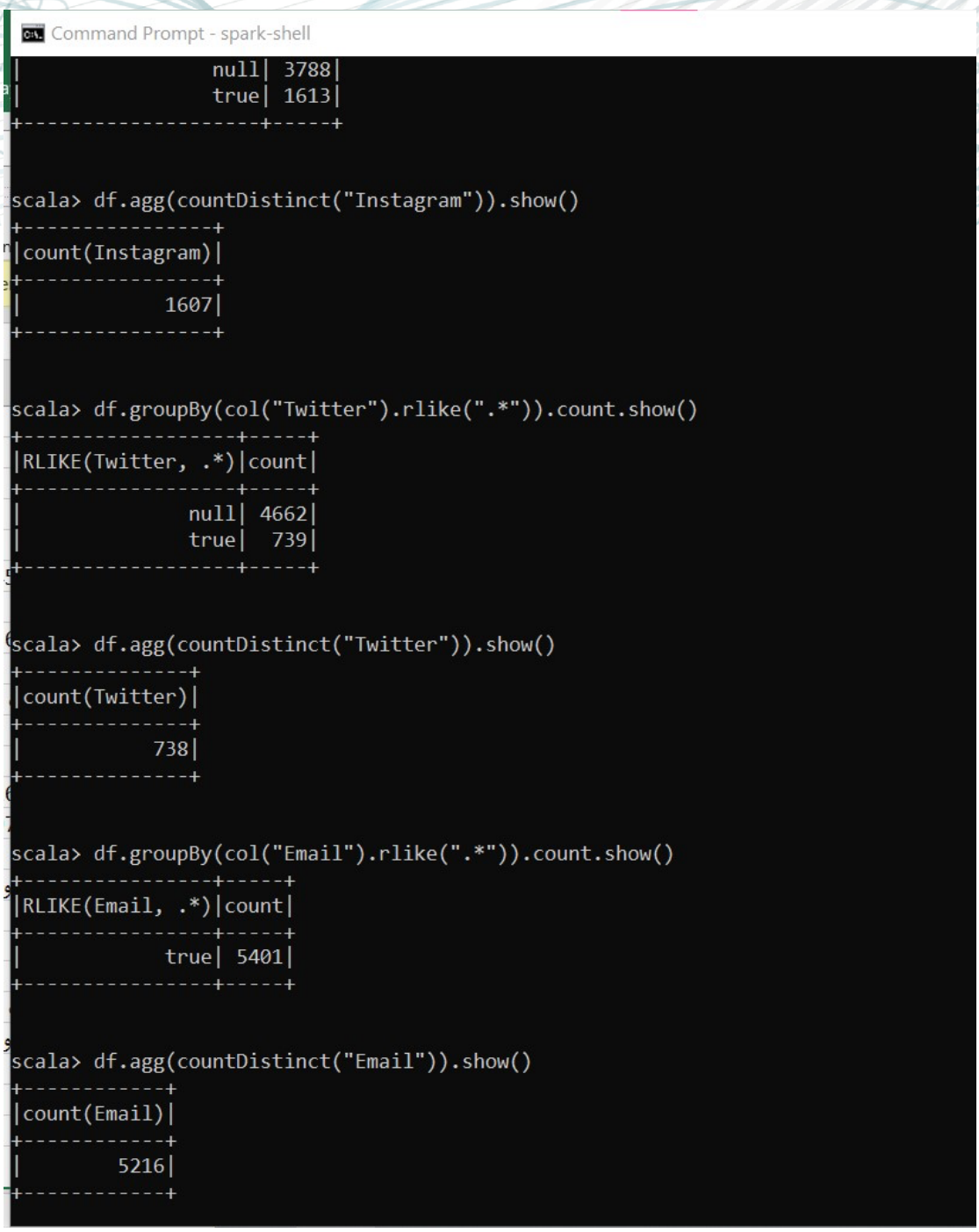
1. Count valid and missing values of Name_ENG column.
2. Count Unique values of Name_ENG column.
3. Count valid and missing values of Name_AR column.
4. Count Unique values of Name_AR column.



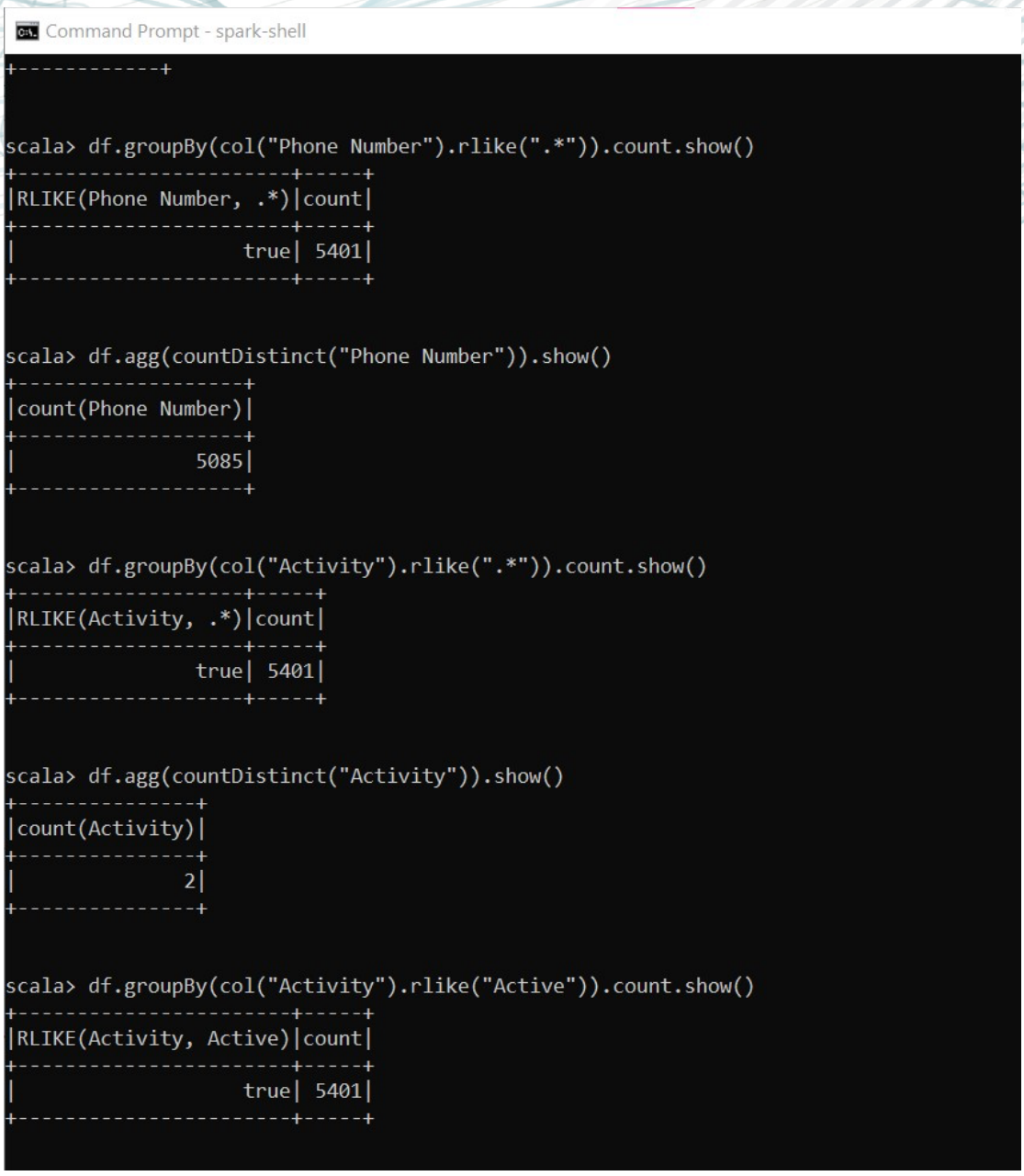
1. Count valid and missing values of CR column.
2. Count Unique values of CR column.
3. Calculate average, maximum and minimum value of CR column.



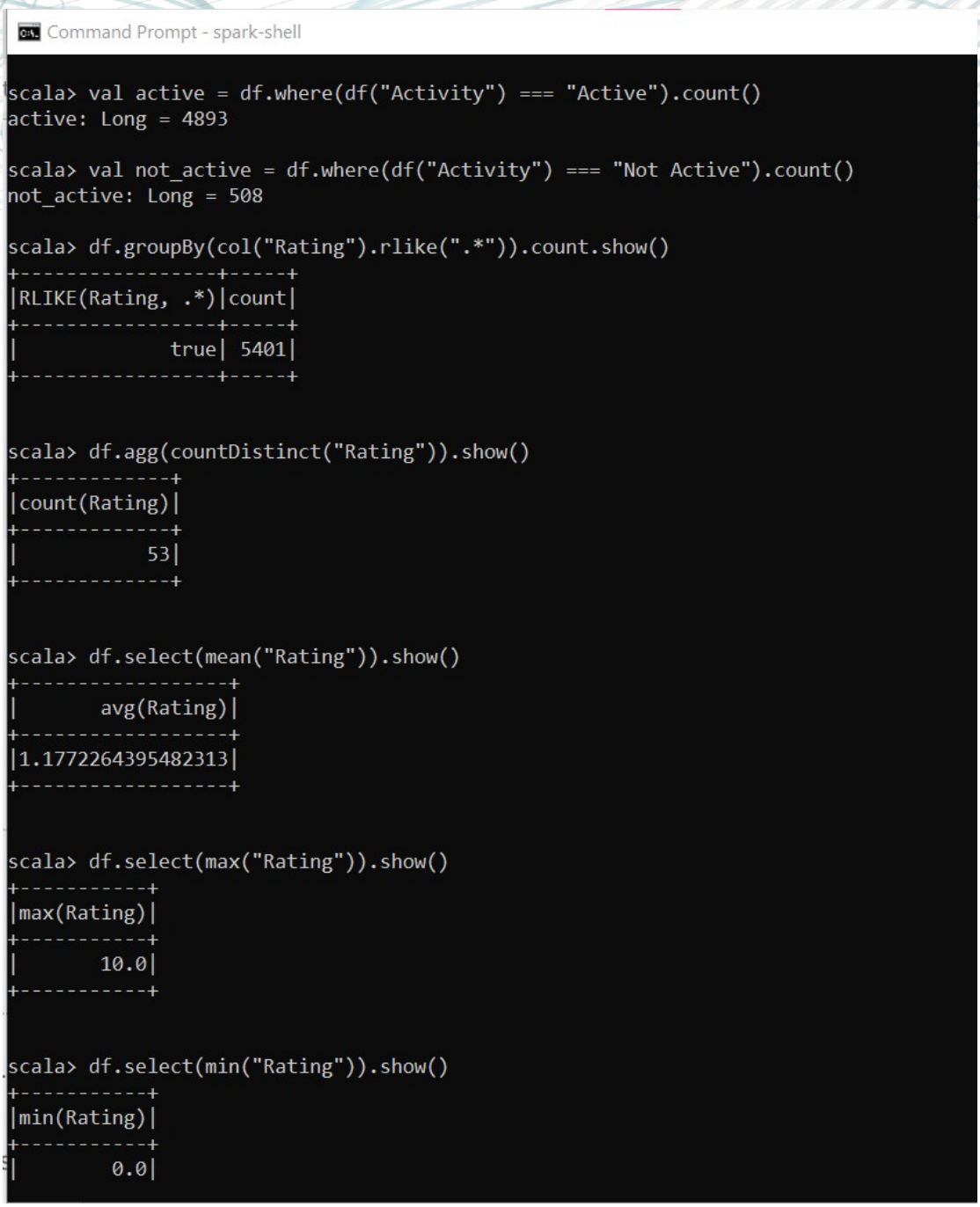
1. Count valid and missing values of Category column.
2. Count Unique values of Category column.
3. Count valid and missing values of Website column.
4. Count Unique values of Website column.
5. Count valid and missing values of Instagram column.
6. Count values of Instagram column.



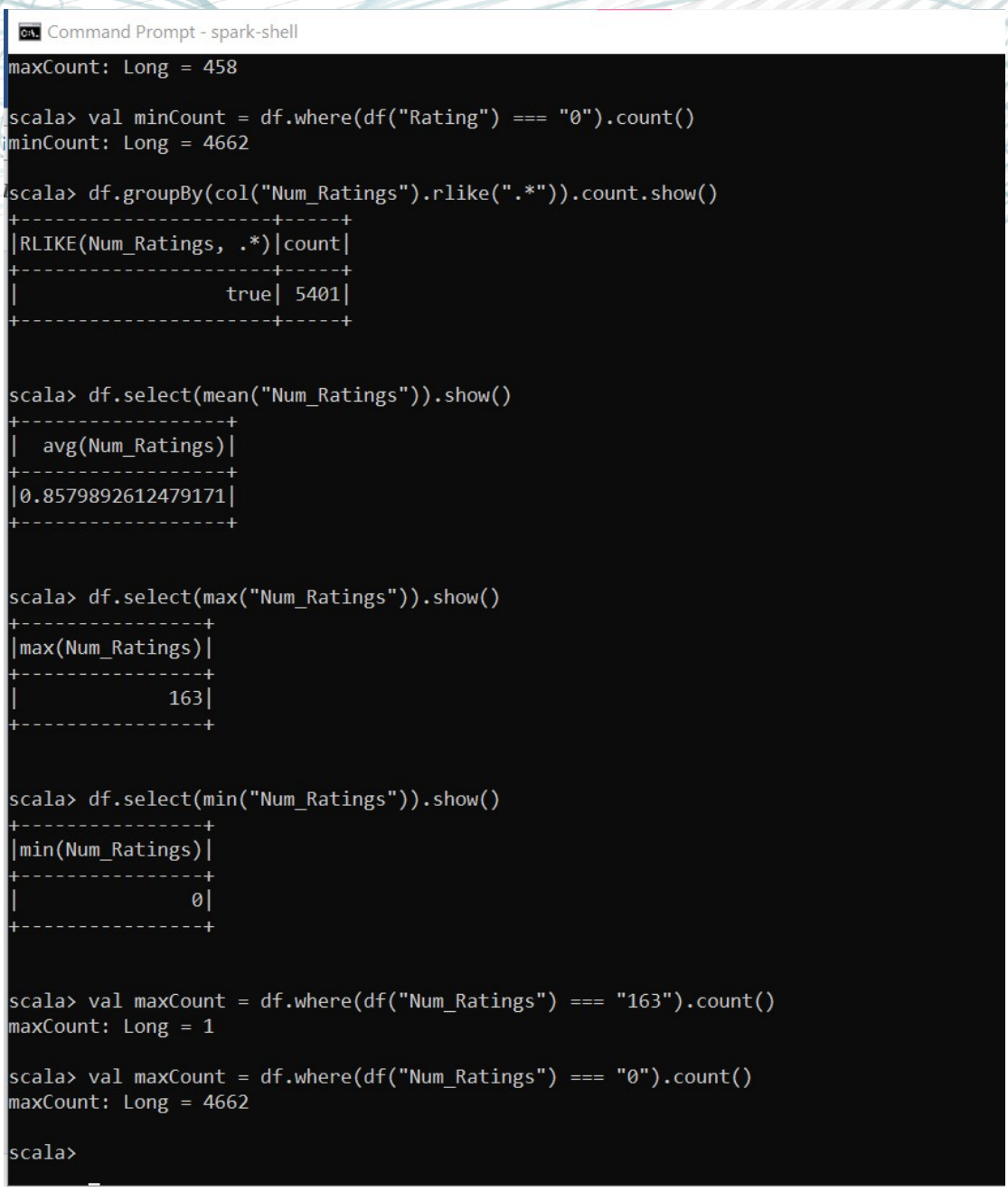
1. Count valid and missing values of Twitter column.
2. Count Unique values of Twitter column.
3. Count valid and missing values of email column.
4. Count values of email column.



1. Count valid and missing values of Phone_Number column.
2. Count Unique values of Phone_Number column.
3. Count valid and missing values of Activity column.
4. Count values of Activity column.



1. Count stores number that are Active.
2. Count store number that are NOT Active.
3. Count valid and missing values of Rating column.
4. Count Unique values of Rating column.
5. Calculate average, maximum and minimum value of Rating column.



1. Count number of stores that have maximum Rating.
2. Count number of stores that have minimum Rating.
3. Count valid and missing values of Num_Rating column
4. Calculate average, maximum and minimum value of Num_Rating column.
5. Count number of stores that have maximum Num_Rating.
6. Count number of stores that have minimum Num_Rating.

Data file used for analysis

Name_ENG	Name_AR	CR	Category	Website	Instagram	Twitter	Email	Phone Number	Activity	Rating
SpeshBox	سبيش بوكس		ديكور	https://speshbox.com	https://instagram.com/speshboxstore		Speshboxstore@gm	555069413	Active	0
	متجر بيع عبايات راقية		تسويق إلكتروني	https://t.me/[illegible]_rw			Noodplm11@gmail.	531303375	Active	0
	شحن شدات [illegible]		تسويق إلكتروني	https://t.me/PO_jo			Noodplm11@gmail.	591203743	Active	2
	تسويق سوقات من المملكة العربية		أخرى	https://t.me/joinchat/AITN3ewGOd2Z0Ey			G7beh@gmail.com	530784921	Not Active	0
Business Hub	شركة ملتقى الأعمال الإعلامية	4650545299	خدمات تعليمية	https://businesshub.com.sa/			info@businesshub.cc	538313518	Active	0
	بيع حسابات سناب دعم امتلاك سناب		تسويق إلكتروني	https://t.me/NP_FP			Noodplm11@gmail.	597213210	Active	0
Visioner Vibes	فيجنر فايبز	2051236409	الحرف والصناعات اليدوية	https://visionervibes.com/			khalloodz1@icloud.c	538320879	Active	0
For sell account snap	بيع حسابات سناب دعم امتلاك		تسويق إلكتروني	https://t.me/VR_R7			Noodplm11@gmail.	530784921	Active	0
	بيع عمل [illegible]		الحرف والصناعات اليدوية	https://salla.sa/noufiana/			noufiana@gmail.cor	535284374	Active	0
	[illegible]		خدمات تعليمية	https://twitter.com/AS3D81		https://twitter.com/AS	tuwaigsa@gmail.cor	538612470	Active	0
	[illegible]		تسويق إلكتروني	https://t.me/TX_XW			ghaithhata@gmail.c	538127030	Active	0
M H Store for decora	متجر ام اتش	2055128943	ديكور	https://salla.sa/[illegible]-sa			mmh.usa@gmail.co	535369055	Active	0
Scientific acuj	أوج العلمية	4030397419	خدمات تعليمية	https://acujco.com/			esaa.hamdan@gma	587927551	Active	0
Atheer Alwasel for T	[illegible]		أخرى	Http://atheeralwasel.com			alwasel.atheer@gm	555186844	Active	0
MyJoey Store	متجر ماي جوي		إلكترونيات وإكسسوارات	https://myjoey.com			myjoey@hotmail.co	571631272	Active	0
Carfix	كارفكس		السيارات	https://salla.sa/carfix1			N7md4de@gmail.cc	533068826	Active	0
shop selling products	متجر لبيع المنتجات		تسويق إلكتروني	https://ain.com			ain.astore2021@gm	557452635	Active	0
DisneyRA	[illegible]		أخرى	https://salla.sa/disney	https://instagram.com/disney_ra		disney_ra@outlook	593260324	Active	0
LuxCups	أكواب الفخمة		الحرف والصناعات اليدوية	https://lux-cups.com/	https://instagram.com/luxcups_		luxcups1@gmail.cor	558743448	Active	0
luxury	[illegible]		إلكترونيات وإكسسوارات	https://salla.sa/www.cladi.com			cawd.2030@gmail.c	535905412	Active	0
Rodi beauty	رودي بيوتي		أخرى	https://salla.sa/rodi7			P3roxa@gmail.com	532491257	Active	0
Fair	فير		تسويق إلكتروني	https://maroof.sa/addbusiness/AddBusinessWizar			aloonuns2020@ho	545022144	Active	0
Unique101 boutique	متجر يونيك101		تسويق إلكتروني	https://unique-101.com			54lo7@hotmail.com	522772911	Active	0
shopping store	[illegible]		أخرى	https://joud-es.com			joud-es@hotmail.co	549930575	Active	0
Omar Mohammed Al	مصنع عمر محمد العطاس	4030142125	تسويق إلكتروني	https://www.alattassouq.sa			alattashome@windo	552422808	Active	0
Magical Pens Agenc	وكالة الأقلام الساحرة	1010336829	ديكور	https://magical-pens.com/			magicalpens11@gm	558689030	Active	0
Borsq Telecom	[illegible]	1131170272	أخرى	https://borsq.sa/			store@borsq.sa	530333218	Active	0
P9mh	[illegible]		إلكترونيات وإكسسوارات	https://salla.sa/store_b9mh/			P9mh13@gmail.com	549491846	Active	0
Pure Land Store	متجر بيور لاند		الحرف والصناعات اليدوية	https://purelandstore.	https://www.instagram.com/pureland.store/		purelandstore1@gm	535127175	Active	0
ALFAHDI STOREE	متجر الفهدي		تسويق إلكتروني	https://t.me/ALFHDI_STORE			afahdi@hotmail.cor	530633478	Active	9.1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	Active	0

Reference

https://www.kaggle.com/datasets/abdullahalothman/saudi-ecommerce-data-set?resource=download