Business Case

ABC Electronics is an online retailer that sells consumer electronics to US customers. The company portfolio includes laptops, mobile phones, and cameras as well as a wide variety of IT accessories. With over 10 years in business, managers have set new goals for upcoming years that include:

Increase sales by 35% over the next 2 years

Increase customer base by 15% over the next 2 years.


```
%spark2.sql
select Category, sum(sales) as Sales from Salesview
group by category
```

Category	Sales
Cameras	835,900.067
Laptops	733,046.8613
Mobile Phones	703,502.928

Business Case

To achieve these goals, sales managers would like to get insights from the sales performance of each category of the business and customers to drive effective business decision-making and plan sales & marketing campaigns accordingly.

For this analysis, we will be using two datasets:

Customer_Data
Customer_ID (Primary Key)
Customer_Name
Gender
Age
Graduated
Profession
Work_Experience
Spending_Score
Family_Size
Ever_Married

10 columns
2120 records

Sales Data	
Row_ID	Postal_code
Customer_ID (Primary Key)	Region
Order_ID	Product_ID
Order_date	Category
Ship_date	Sub_category
Ship_mode	Product_name
Customer_name	Sales
Segment	Quantity
Country	Discount
City	Profit
State	Order_type

22 columns
793 records

Data Processing Hadoop Tools

External Table in Hive - Customer data

```
CREATE EXTERNAL TABLE IF NOT EXISTS customer_external(
    Customer_ID string,Customer_Name string,Gender string,Age int,Graduated string,Profession string,Work_Experience int,
    Spending_Score string,Family_Size int,Ever_Married string)
    COMMENT 'Data about customers from a public database'
    ROW FORMAT DELIMITED
    FIELDS TERMINATED BY ','
    STORED AS TEXTFILE
    location '/user/maria_dev/customer_dataproject';
```

Dataframe in Zeppelin - Sales data

```
%spark2.spark
val salesdf = (spark.read
.option("header", "true")                    // Use first line as header
.option("inferSchema", "true")                  // Infer schema
.csv("/tmp/sales_datagp.csv"))
```

```
salesdf: org.apache.spark.sql.DataFrame = [Row_ID: int, Order_ID: string ... 20 more fields]
```

Data Processing Hadoop Tools

Temporary View

```
%spark2.spark
salesdf.createOrReplaceTempView("Salesview")
```

Data is ready for analysis

```
%spark2.sql
select * from Salesview
```

FINISHED

Row_ID	Order_ID	Order_date	Ship_date	Ship_mode	Customer_ID	Customer_name	Segment	Country	City	State	Post
1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	4242

```
%spark2.sql

select * from customer_external
```

customer_id	customer_name	gender	age	graduated	profession	work_experience	spending_score	family_size
CG-12520	Claire Gute	F	36	Yes	Engineer	0	Low	1
DV-13045	Darrin Van Huff	M	37	Yes	Healthcare	8	Average	4

Hive to Hbase to Zeppelin Analysis

Creation of External Table

```
CREATE EXTERNAL TABLE IF NOT EXISTS customer_external(
    Customer_ID string,Customer_Name string,Gender string,Age int,Graduated string,Profession string,Work_Experience int,Spending_Score string,
    Family_Size int,Ever_Married string)
    ROW FORMAT DELIMITED
    FIELDS TERMINATED BY ','
    STORED AS TEXTFILE
    LOCATION '/user/maria_dev/Grp_project';
```

Creation of hive internal table

```
CREATE TABLE IF NOT EXISTS customer_orc(
    Customer_ID string,Customer_Name string,Gender string,Age int,Graduated string,Profession string,Work_Experience int,Spending_Score string,
    Family_Size int,Ever_Married string)
    STORED AS ORC;
```

Hive to Hbase to Zeppelin Analysis

Loaded from External into Internal ORC

```
INSERT INTO TABLE customer_orc SELECT * FROM customer_external
```

We then logged into hbase and created hbase table

```
hbase(main):006:0> create 'customer', 'details'
0 row(s) in 1.2370 seconds

=> Hbase::Table - customer
hbase(main):007:0>
```

Created table in hive that maps directly to the hbase

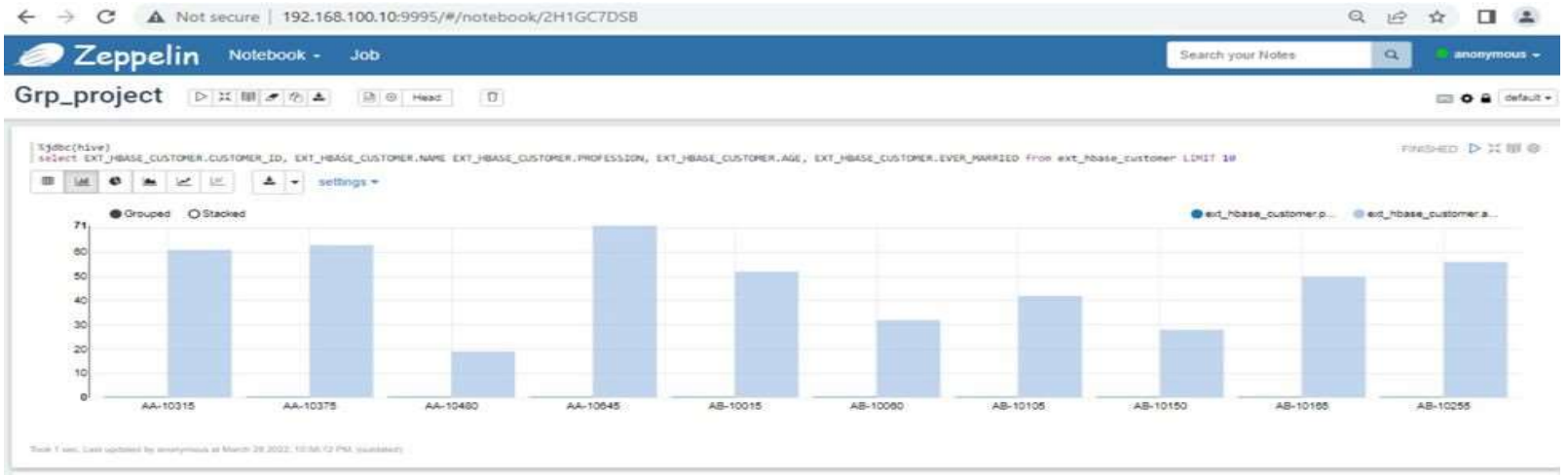
```
CREATE EXTERNAL TABLE ext_hbase_customer (Customer_ID string,Customer_Name string,Gender string,Age int,Graduated string,Profession string,Work_
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,details:Customer_Name,details:Gender,details:Age,details:Graduated,details:Profession,deta
TBLPROPERTIES("hbase.table.name" = "customer","hbase.mapred.output.outputtable" = "customer")
```

Hive to Hbase to Zeppelin Analysis

From hive we inserted the data into hbase table

```
INSERT INTO TABLE ext_hbase_customer SELECT * FROM customer_orc
```

After that we checked the data in zeppelin



Business Analysis

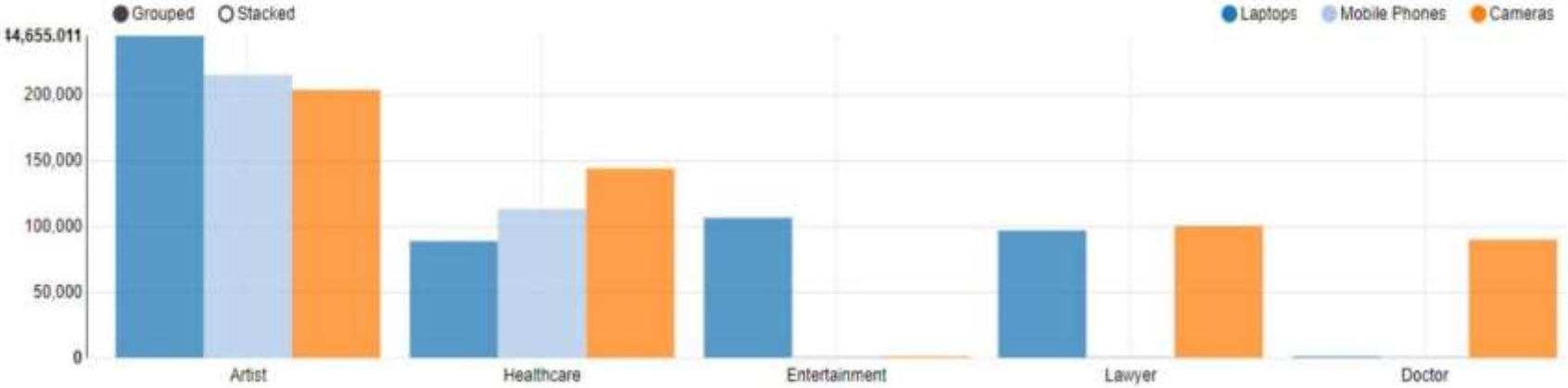
1. Sale of each product by the profession of customer

Business Insight - To get to know about which product has higher demand in which profession so that company can avail options with suitable monthly payment plans or discount rates to adjust the products profit margin.

```
%spark2.sql
select s.category, sum(s.sales) as sales, c.profession
from Salesview s join customer_external c
on s.customer_ID = c.customer_id
group by s.category, c.profession
order by Sales desc
limit 10
```

FINISHED

settings



Business Analysis

2. Total number of orders placed by customers aged - less than 25 , between 25 to 50, and above 50 in the year 2019.

Business Insight - Sales manager will get an insight about which age category is buying which product more of their company to decide their marketing strategy accordingly. For age group < 25 and age group 25 to 50, marketing can be done through social media.

A**.** For age < 25 years

```
%spark2.sql
select count(*) as Total_Count_of_Products,
       s.category
       from SalesData s inner join customer_external c
       on (s.Customer_Id = c.customer_id AND c.age < 25 AND order_date between '01/01/2019' and '12/31/2019')
       group by s.category
       order by s.category
```

FINISHED

settings

Took 2 sec. Last updated by anonymous at March 31 2022, 12:19:20 AM. (outdated)

Business Analysis

2. Total number of orders placed by customers aged - less than 25 , between 25 to 50, and above 50 in the year 2019.

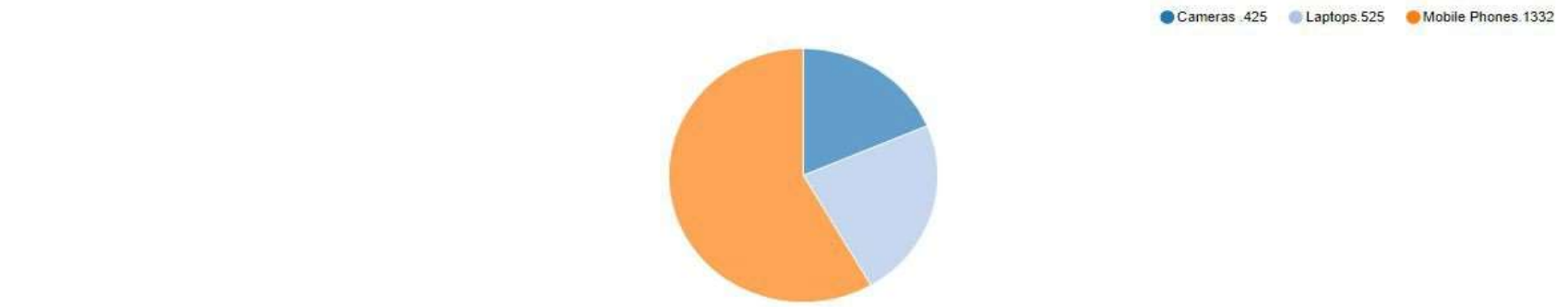
Business Insight - Sales manager will get an insight about which age category is buying which product more of their company to decide their marketing strategy accordingly. For age group < 25 and age group 25 to 50, marketing can be done through social media.

B. For age between 25 to 50

```
%spark2.sql
select count(*) as Total_Count_of_Product,
       s.category
       from SalesData s inner join customer_external c
       on s.Customer_Id = c.customer_id
       where ((c.age >= 25 AND c.age <= 50)  AND order_date between '01/01/2019' AND '12/31/2019')
       group by s.category
       order by s.category
```

FINISHED

settings



Took 2 sec. Last updated by anonymous at April 03 2022, 12:53:22 PM.

Business Analysis

2. Total number of orders placed by customers aged - less than 25 , between 25 to 50, and above 50 in the year 2019.

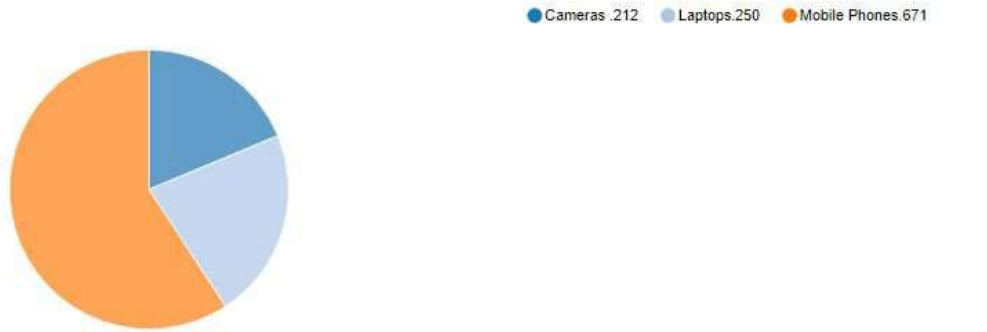
Business Insight - Sales manager will get an insight about which age category is buying which product more of their company to decide their marketing strategy accordingly. For age group > 50, marketing can be done through newspaper, magazine, tv commercials.

C. For age above 50

```
%spark2.sql
select count(*) as Total_Count_of_Product,
       s.category
       from SalesData s inner join customer_external c
       on s.Customer_Id = c.customer_id
       where ( c.age > 50 AND order_date between '01/01/2019' and '12/31/2019')
       group by s.category
       order by s.category
```

FINISHED

settings



Took 2 sec. Last updated by anonymous at March 31 2022, 12:20:50 AM. (outdated)

Business Analysis

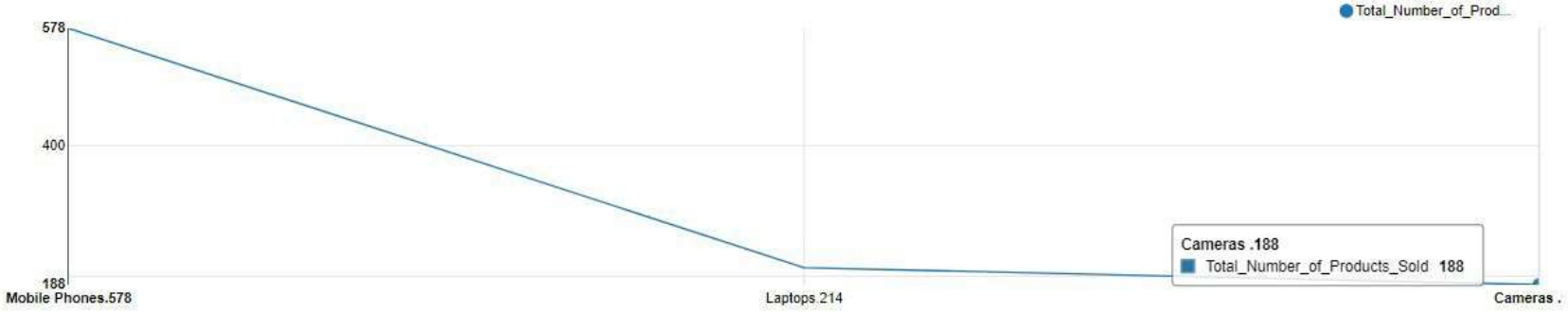
3. **Product getting sold the least to the customers for age group less than 25.**

Business Insight - Sales manager will analyse which product is getting sold least so as to have any specific discounts on that product to make them sale and make profit out of it.

```
%spark2.sql
select s.category,
       count(*) as Total_Number_of_Products_Sold
       from SalesData s inner join customer_external c
       on s.Customer_Id = c.customer_id
       where c.age < 25
       group by s.category
       order by s.category desc
```

FINISHED

settings



Took 1 sec. Last updated by anonymous at March 31 2022, 12:25:36 AM. (outdated)

Business Analysis

4. **Purchased product information of all customers who has spending score = low**

Business Insight - Sales manager can think of increasing customers count from low to medium/high spending score so as to make profit out of it by offering multiple credit plans.

```
%spark2.sql
select s.*
      from SalesData s inner join customer_external c
      on s.Customer_Id = c.customer_id
      where c.spending_score = 'Low'
```

FINISHED

Row_ID	Order_ID	Order_date	Ship_date	Ship_mode	Customer_ID	Customer_name	Segment	Country	City	State	Postal Code	Region
1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South
2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South
4	US-2019-108966	10/11/2019	10/18/2019	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South
5	US-2019-108966	10/11/2019	10/18/2019	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South
13	CA-2017-114412	4/15/2017	4/20/2017	Standard	AA-10480	Andrew Allen	Consumer	United	Concord	North Carolina	28027	South

Output is truncated to 102400 bytes. Learn more about ZEPPELIN_INTERPRETER_OUTPUT_LIMIT

Business Analysis

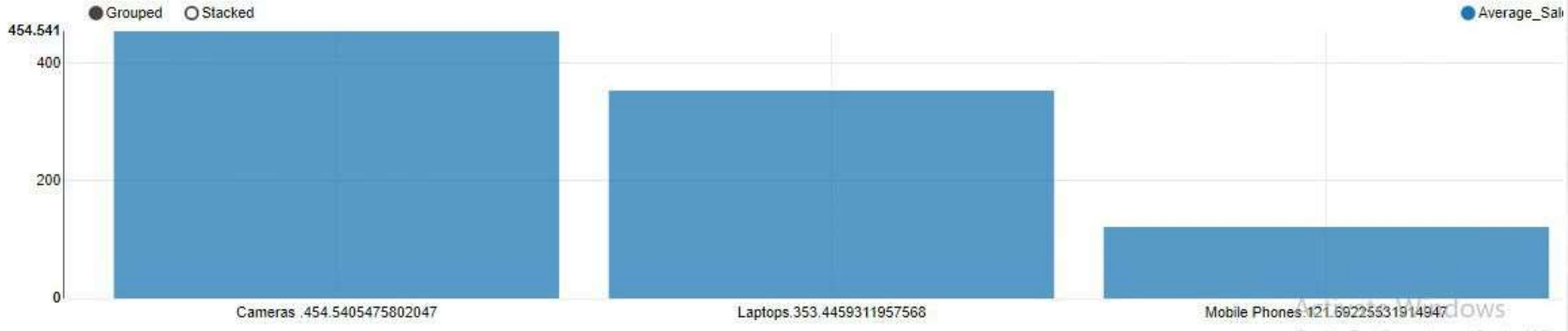
5. Average sale of each product category.

Business Insight - To get to know which product category is at which position in terms of sale. Management can decide the inventory for respective product in order to minimize the holding cost which can eventually contribute towards more profit.

```
%spark2.sql
select category,
       avg(Sales) as Average_Sale
       from SalesData
       where category IN('Cameras ','Laptops','Mobile Phones')
       group by category
       order by category
```

FINISHED

settings



Business Analysis

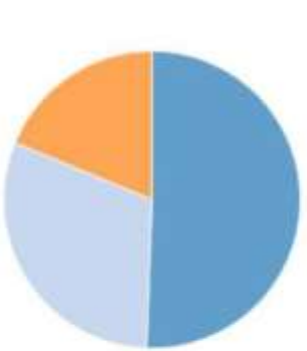
6. Sales by customer segment

Business Insight - Products are most appealing to the consumer segment. Marketing strategies can be built from this insight.

```
%spark2.sql
select Segment, sum(sales) as Total_Sales from sales_data_view
group by Segment
order by Total_Sales desc
```

FINISHED

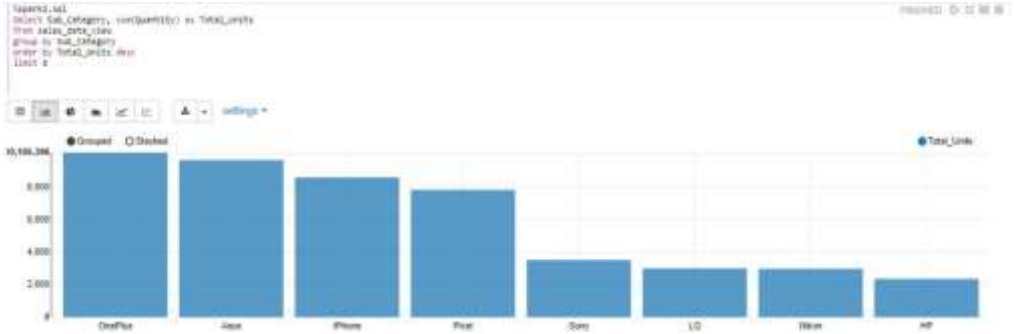
settings



Took 2 sec. Last updated by anonymous at April 01 2022, 10:03:30 PM. (outdated)

Business Analysis

7. Most and least Popular Brands sold - DC storage operations can be strategized using this insight

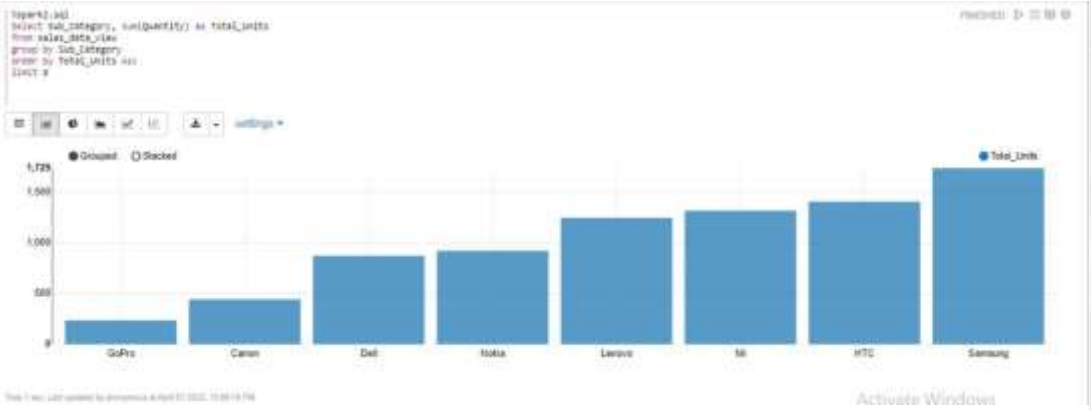


Most Popular Brands sold

Business Insight - OnePlus, Asus, iPhone, etc. are fast moving items and can be ordered from vendor in pallet amounts and stored closer to the docks

Least Popular Brands sold

Business Insight - GoPro and Canon have negligible sales. Marketing can focus on increasing sales for these brands or management can decide to terminate contracts with these brands to avoid relatively large fixed contractual costs

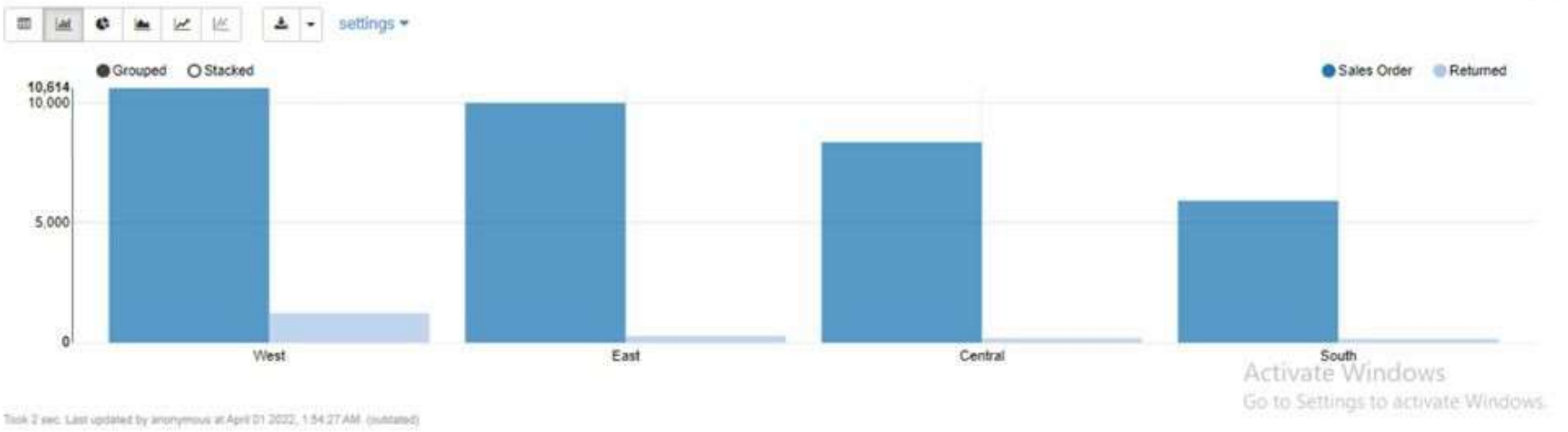


Business Analysis

8. Return vs Sales by Region (Total units)

Business Insight - The Western Region has the maximum returns. This would indicate proper QA measures need to be set up to overcome this inefficiency.

```
%spark2.sql
Select Region, Order_type, sum(Quantity) as Total_Units
from salesview
where Order_type = "Sales Order" or Order_type = "Returned"
group by Order_type, Region
order by Total_Units desc
```



Business Analysis

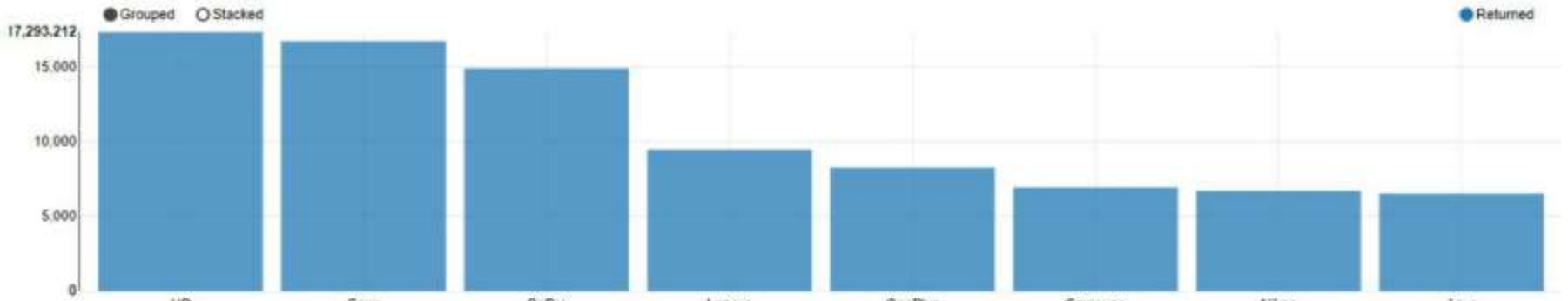
9. Brands with maximum returns

Business Insight - This Insight can drive re-negotiation contracts with the brand vendors to either improve their quality or reduce the prices.

```
%spark2.sql
Select Order_type, Sub_Category, sum(Sales) as Total_Sales
from sales_data_view
where Order_type = "Returned"
group by Order_type, Sub_Category
order by Total_Sales desc
limit 8
```

FINISHED

settings



Conclusion

This analysis can help the business drive various decisions and strategies through these insights:

- Understanding the customer base
- Procurement strategies
- Marketing strategies
- Negotiations with vendors
- Strategies to increase market share