

# EAS 596 - CYBERSECURITY ANALYTICS

## ASSIGNMENT 02

### GROUP- 5:

Ayyathurai Jaya, Raghavi Preeti

Heeb, Collin

Kandasamy, Kathiresan

Patnayakuni, Sharada




Pindi, Naren

1) Utilize a sparkline to indicate the count of sales over the entire time-period for each location.

### QUERY:

```
index="store_sales"
| timechart span=1d count by Branch
| foreach * [eval <<FIELD>>=coalesce('<<FIELD>>', 0)]
| untable _time Branch count
| stats sparkline(count) as Sales_Trend, count as Total_Sales by Branch
```

### OUTPUT:

Branch	Sales_Trend	Total_Sales
A		1999
B		1999
C		1999

2) Look at the top 5 product lines and give the total sales for each along with the amount of the most expensive item.

### QUERY:

```
index=main sourcetype=pharmasalescsv earliest=0
| eval fields=split(_raw, ",")
| eval ProductLine=mindex(fields,5), Total=tonumber(mindex(fields,9)), Price=tonumber(mindex(fields,6))
| stats sum(Total) as total_sales, max(Price) as most_expensive_item by ProductLine
| sort - total_sales
| head 5
```

### OUTPUT:

ProductLine	total_sales	most_expensive_item
6	336771	5
7	334762	5
5	322644	5
4	320628	9
9	310570	6

### 3) Which items sell the most overall? Per invoice?

#### (i) OVERALL

##### QUERY:

```
index=main sourcetype=pharmasalescsv earliest=0
| eval fields=split(_raw, ",")
| eval ProductLine=mvindex(fields,5), Quantity=tonumber(mvindex(fields,7))
| stats sum(Quantity) as total_sold by ProductLine
| sort - total_sold
| head 10
```

##### OUTPUT:

ProductLine	total_sold
12	590
11	888
10	801.33333333
9	797.29166667
8	859
7	852.25
6	945
5	882
4	833
3	725.875

#### (ii) PER INVOICE:

##### QUERY:

```
index=main sourcetype=pharmasalescsv earliest=0
| eval fields=split(_raw, ",")
| eval ProductLine=mvindex(fields,5), InvoiceID=mvindex(fields,0), Quantity=tonumber(mvindex(fields,7))
| stats sum(Quantity) as qty_per_invoice by InvoiceID, ProductLine
| stats avg(qty_per_invoice) as avg_per_invoice by ProductLine
| sort - avg_per_invoice
| head 10
```

##### OUTPUT:

ProductLine	avg_per_invoice
43	13
31	11.5
28.33333333	21
23.33333333	12.5
18.75	13.75
17.66666667	12
14.58333333	11.25
11.25	10
10.83333333	11.8750000015
2.6	11

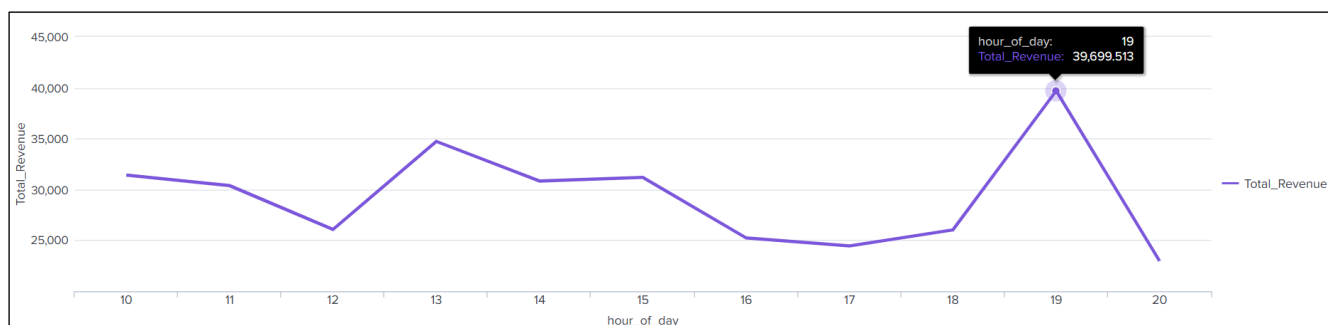
4) Do an analysis on the time of day. Is there any indication that any time of the day is most profitable?

QUERY:

```
index="store_sales"
| eval hour_of_day = strftime(strptime(Time, "%H:%M"), "%H")
| stats sum(Total) as Total_Revenue by hour_of_day
| sort hour_of_day
```

OUTPUT:

hour_of_day	Total_Revenue
10	31421.4810
11	30377.3295
12	26065.8825
13	34723.2270
14	30828.3990
15	31179.5085
16	25226.3235
17	24445.2180
18	26030.3400
19	39699.5130
20	22969.5270



Based on the time-of-day analysis, the highest total revenue was observed around **7:00 PM (hour 19)**, indicating that the evening hours are the most profitable period for the store.

5) Break down, per each branch, the differences in spending between men and women. Explain why you selected the fields that you did and why this query would be sufficient for management to understanding spending habits based on gender.

QUERY:

```
index="store_sales"
| chart sum(Total) over Branch by Gender
| eval Total_Spending = Male + Female
```

OUTPUT:

Branch	Female	Male	Total_Spending
A	53269.1670	52931.2035	106200.3705
B	52928.2950	53269.3770	106197.6720
C	61685.4630	48883.2435	110568.7065

We selected the **Branch** field to isolate data by physical store location, and the **Gender** field to analyze shopping behavior differences between male and female customers. This analysis is important for management because it helps identify which gender spends more at each branch.

For example, in **Branch C**, if females are spending more than males, management might consider promoting **female-oriented products or offers** to further increase engagement and sales. This approach can help create **gender-specific strategies** across all branches, ultimately boosting overall revenue by catering to the spending habits of different customer groups.

6) Come up with a separate search that you believe will be important for management of this organization to understand. Explain your reasoning.

QUERY:

```
index="store_sales"
| stats sum(Total) as Revenue by Branch "Product line"
| sort -Revenue
```

OUTPUT:

Branch	Product line	Revenue
C	Food and beverages	23766.8550
A	Home and lifestyle	22417.1955
C	Fashion accessories	21560.0700
B	Sports and travel	19988.1990
B	Health and beauty	19980.6600
A	Sports and travel	19372.6995
C	Electronic accessories	18968.9745
A	Electronic accessories	18317.1135

This query helps management understand which **product lines** are generating the **most revenue in each branch**. It combines product performance with branch performance, providing insights into regional customer preferences.

For example, if the **"Food and beverages"** category performs well in **Branch C** but poorly in **Branch A**, management can adjust marketing strategies, inventory levels, or even store layout to better match customer demand.