📊 Fraud Detection with Splunk Enterprise — Advanced SOC Dashboard Project

📝 **Date:** July 12, 2025
🧠 **Analyst:** Aarush Nepali
🖥️ **Environment:** Splunk Enterprise on Kali Linux
🗂️ **Dataset:** `prepared_data.xlsx` → `prepared_data.csv`
🎯 **Goal:** Identify and visualize fraudulent transactions based on behavioral and demographic indicators using real-world Splunk ingestion, SPL queries, and dashboarding.

---

1. Executive Summary

This project demonstrates hands-on expertise in building a fraud detection workflow using **Splunk Enterprise** for real-time pattern recognition and visualization. Working on Kali Linux, I transformed raw transactional data into actionable intelligence by identifying the **who, what, when, and where** of fraudulent behaviors. The analysis highlights high-risk demographics, targeted merchants, and time-based fraud trends — offering direct value for security operations teams and fraud monitoring programs.

🔍 2. Problem Overview

A structured transaction dataset, potentially containing fraudulent records, required ingestion and analysis in Splunk. The goal was to answer:

- Which **demographic groups** commit fraud most frequently?
- What **purchase categories** and **merchants** are most exploited?
- Are there **seasonal or temporal trends** in fraud activity?
- What is the **geographic origin** of the majority of incidents?

---

3. Data Ingestion Workflow

Original File:
`/home/kali/Desktop/prepared_data.xlsx`

Issue Encountered:

Uploading `.xlsx` into Splunk produced **corrupted hex characters** (e.g., `\xC4\xB1...`) due to binary incompatibility.

🔧 **Resolution:**

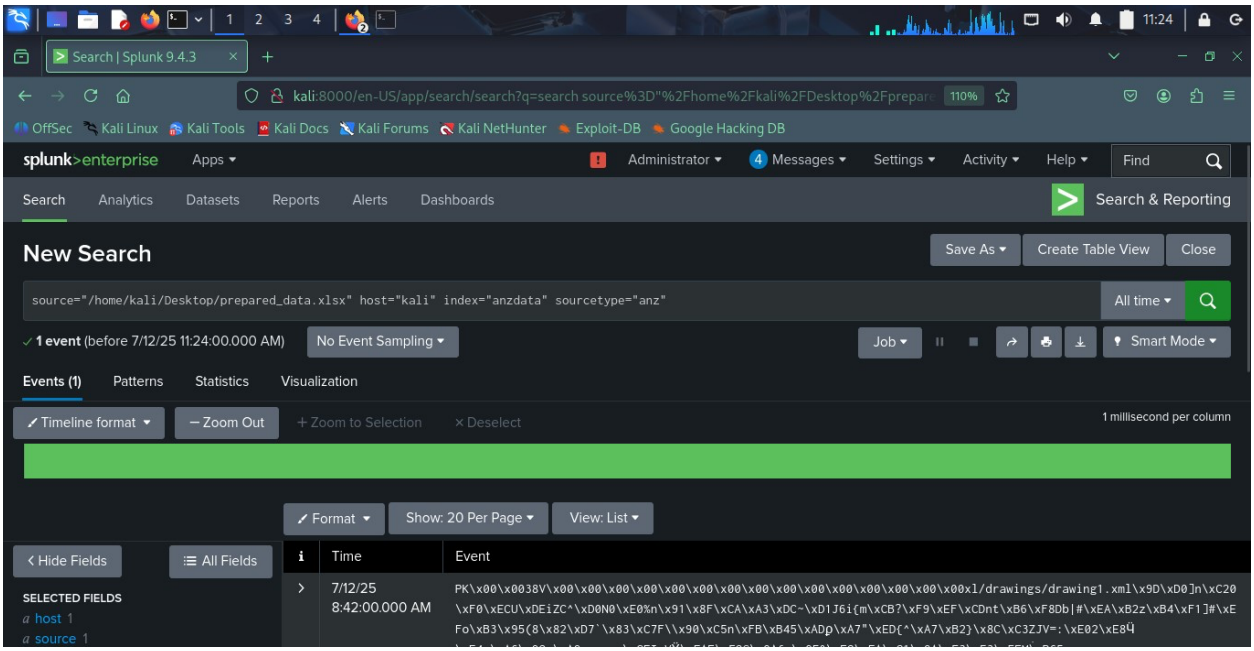
Converted `.xlsx` to `.csv` using:

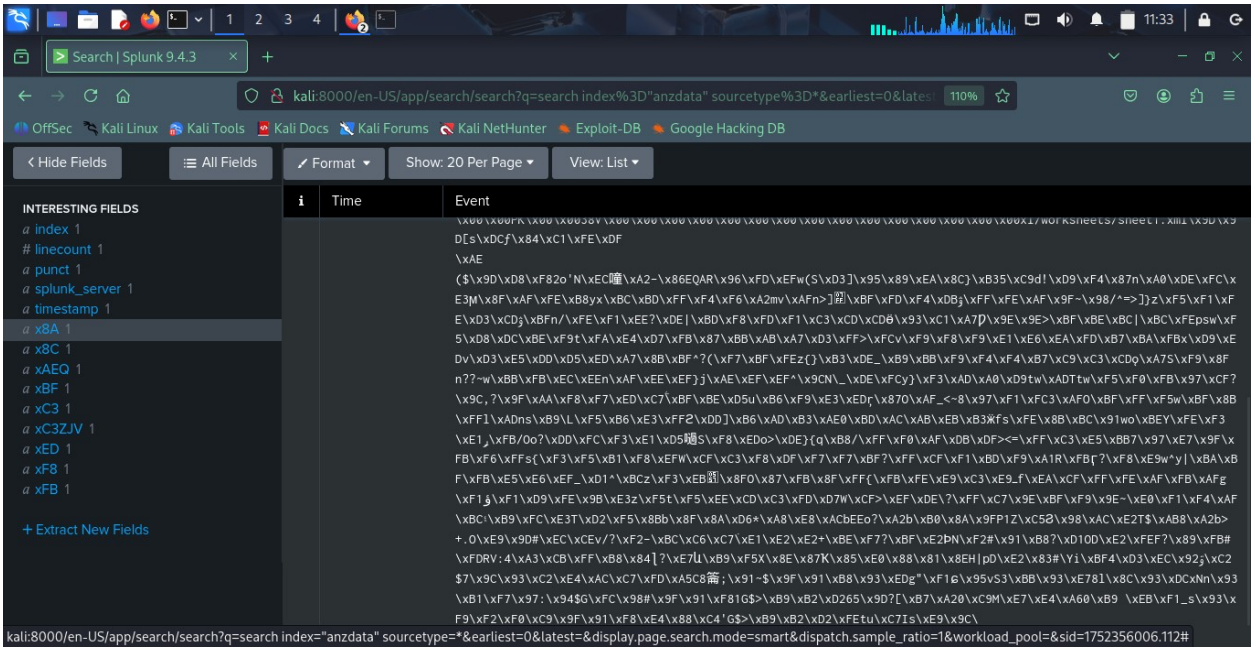
```
bash ssconvert prepared_data.xlsx
prepared_data.csv
```

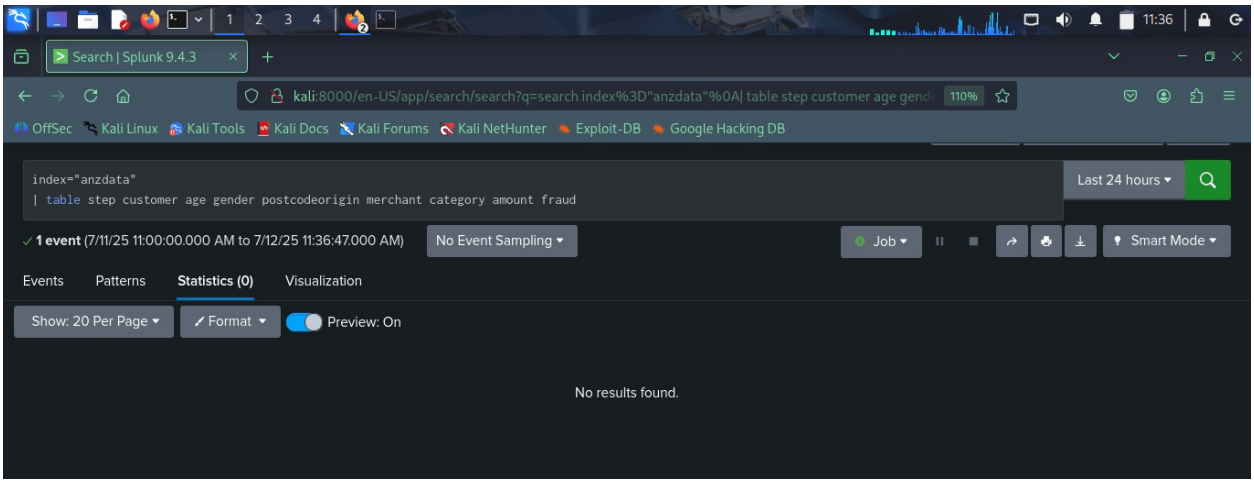
Ingested into Splunk via CLI:

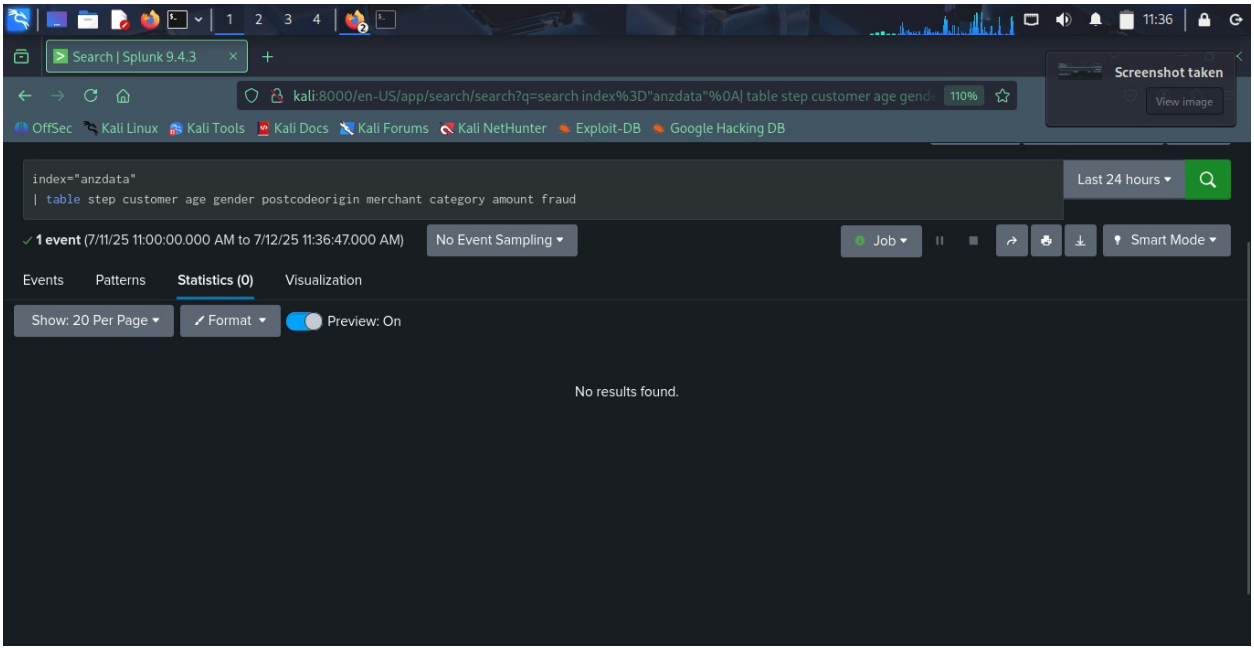
```
bash
/opt/splunk/bin/splunk add oneshot /home/kali/Desktop/prepared_data.csv index
anzdata -sourcetype csv
```

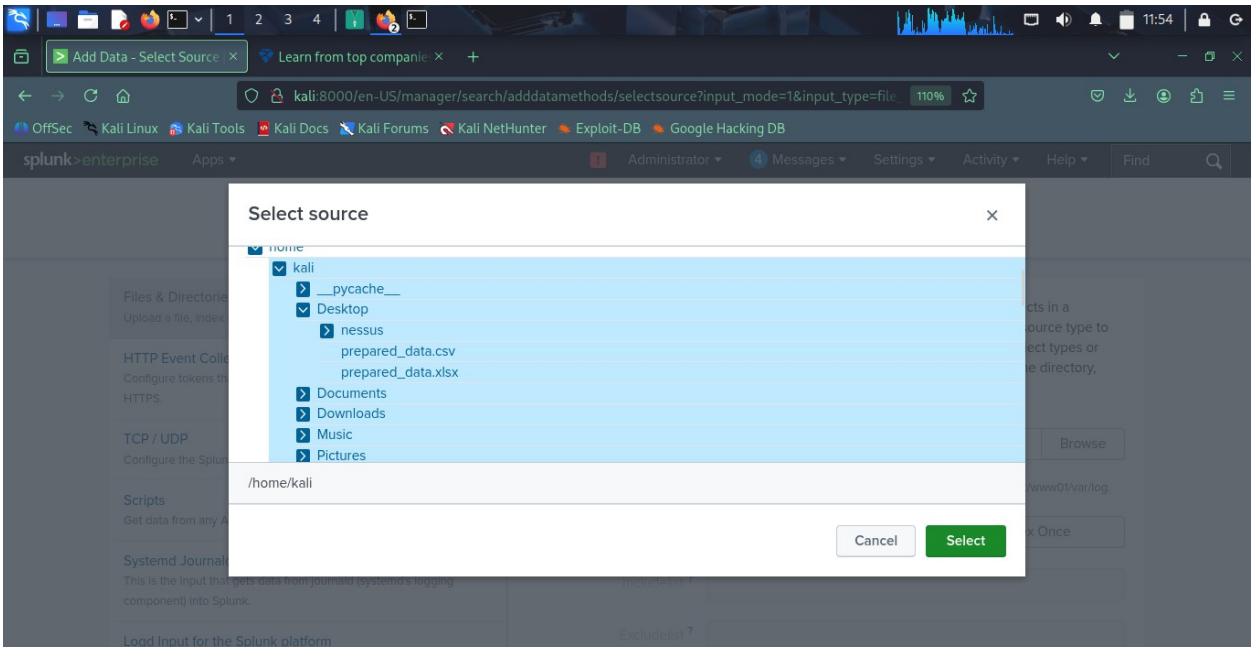
- **Index:** `anzdata`
- **Sourcetype:** `csv`

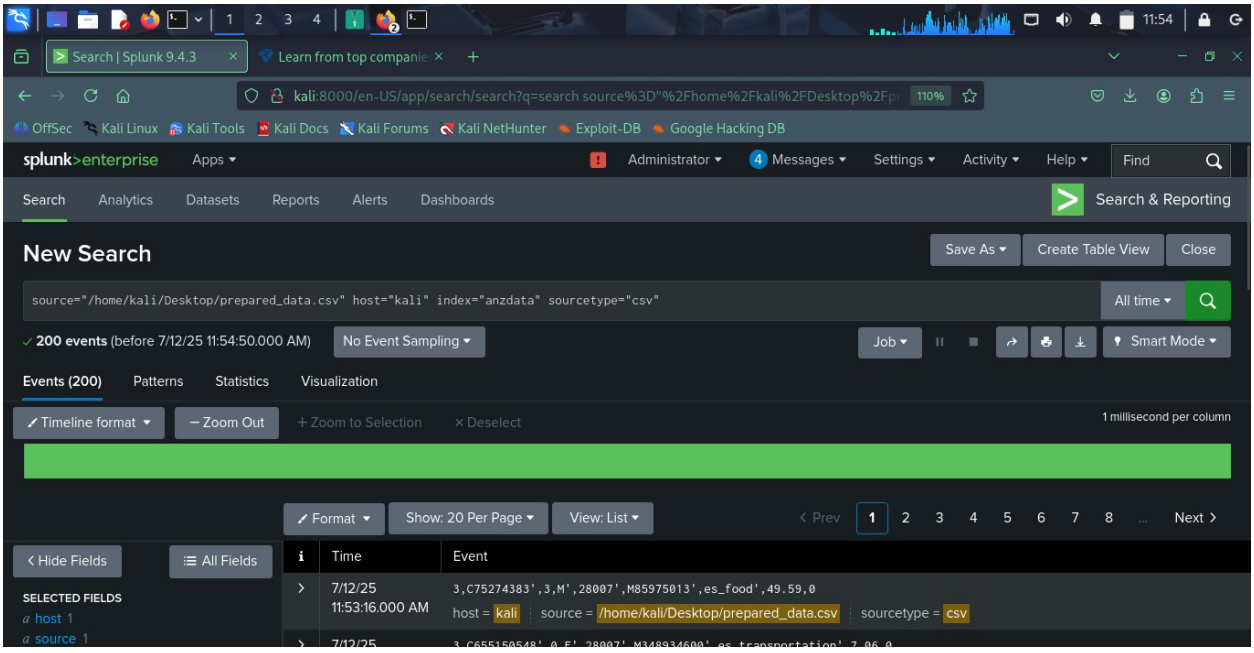

Search | Splunk 9.4.3
splunk>enterprise
Search & Reporting
New Search
Save As
Create Table View
Close
source="/home/kali/Desktop/prepared_data.xlsx" host="kali" index="anzdata" sourcetype="anz"
All time
1 event (before 7/12/25 11:24:00.000 AM)
No Event Sampling
Smart Mode
Events (1)
Patterns
Statistics
Visualization
Timeline format
Zoom Out
Zoom to Selection
Deselect
1 millisecond per column
Format
Show: 20 Per Page
View: List
Hide Fields
All Fields
SELECTED FIELDS
host 1
source 1
Time
Event
7/12/25 8:42:00.000 AM

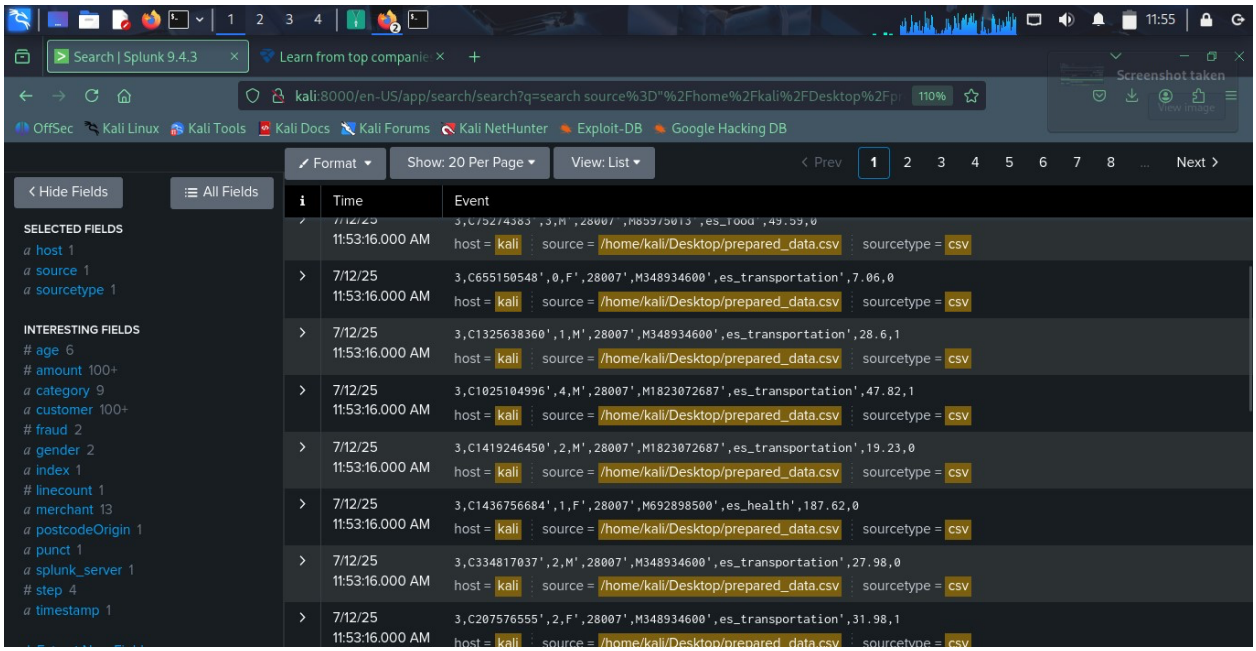

Search | Splunk 9.4.3
Hide Fields
All Fields
Format
Show: 20 Per Page
View: List
INTERESTING FIELDS
index 1
linecount 1
punct 1
splunk_server 1
timestamp 1
x8A 1
x8C 1
xAEQ 1
xBF 1
xC3 1
xC3ZJV 1
xED 1
xF8 1
xFB 1
Extract New Fields
Time
Event

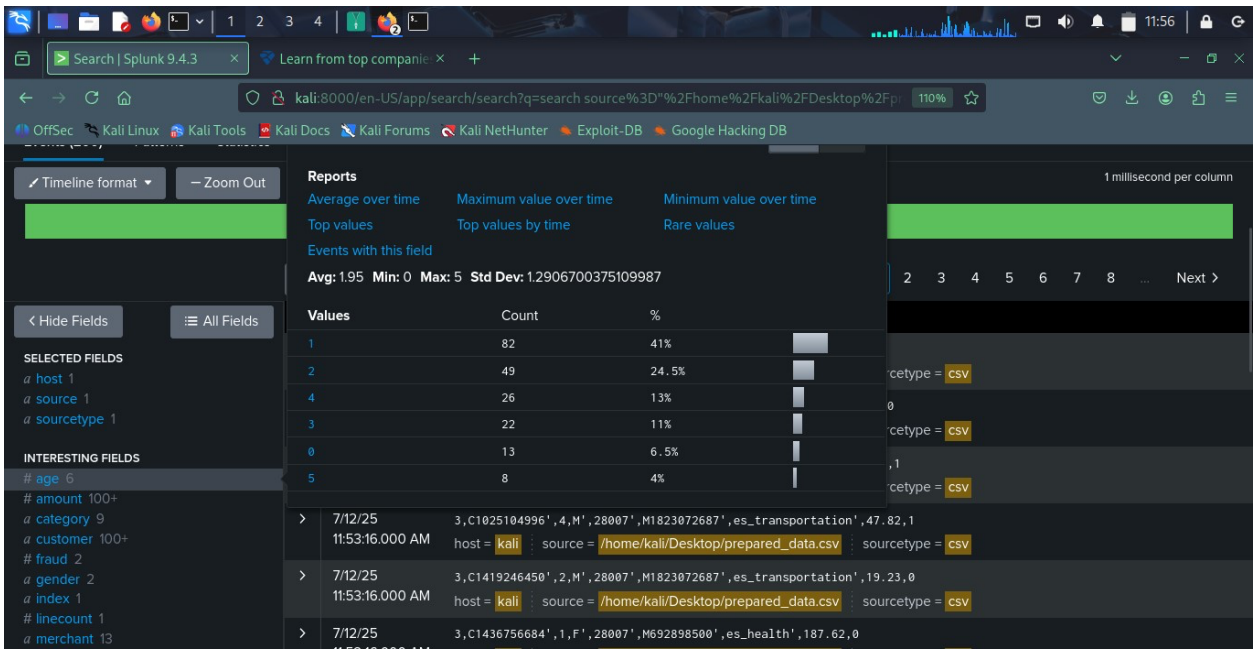

Search | Splunk 9.4.3
kali:8000/en-US/app/search/search?q=search index%3D"anzdata"%0A| table step customer age gend
110%
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
index="anzdata"
| table step customer age gender postcodeorigin merchant category amount fraud
Last 24 hours
1 event (7/11/25 11:00:00.000 AM to 7/12/25 11:36:47.000 AM)
No Event Sampling
Job
Smart Mode
Events
Patterns
Statistics (0)
Visualization
Show: 20 Per Page
Format
Preview: On
No results found.


Search | Splunk 9.4.3
kali:8000/en-US/app/search/search?q=search index%3D"anzdata"%0A| table step customer age gend
110%
Screenshot taken
View image
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
index="anzdata"
| table step customer age gender postcodeorigin merchant category amount fraud
Last 24 hours
1 event (7/11/25 11:00:00.000 AM to 7/12/25 11:36:47.000 AM)
No Event Sampling
Job
Smart Mode
Events
Patterns
Statistics (0)
Visualization
Show: 20 Per Page
Format
Preview: On
No results found.


Add Data - Select Source
Learn from top companie
kali:8000/en-US/manager/search/adddatamethods/selectsource?input_mode=1&input_type=file
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Select source
kali
__pycache__
Desktop
nessus
prepared_data.csv
prepared_data.xlsx
Documents
Downloads
Music
Pictures
/home/kali
Cancel
Select


Search | Splunk 9.4.3
Learn from top companie
kali:8000/en-US/app/search/search?q=search source%3D"%2Fhome%2Fkali%2FDesktop%2Fp
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Search
Analytics
Datasets
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Create Table View
Close
source="/home/kali/Desktop/prepared_data.csv" host="kali" index="anzdata" sourcetype="csv"
All time
200 events (before 7/12/25 11:54:50.000 AM)
No Event Sampling
Job
Smart Mode
Events (200)
Patterns
Statistics
Visualization
Timeline format
Zoom Out
Zoom to Selection
Deselect
1 millisecond per column
Format
Show: 20 Per Page
View: List
Prev
1
2
3
4
5
6
7
8
Next
Hide Fields
All Fields
SELECTED FIELDS
host 1
source 1
Time
Event
7/12/25
11:53:16.000 AM
3,C75274383',3,M',28007',M85975013',es_food',49.59,0
host = kali
source = /home/kali/Desktop/prepared_data.csv
sourcetype = csv



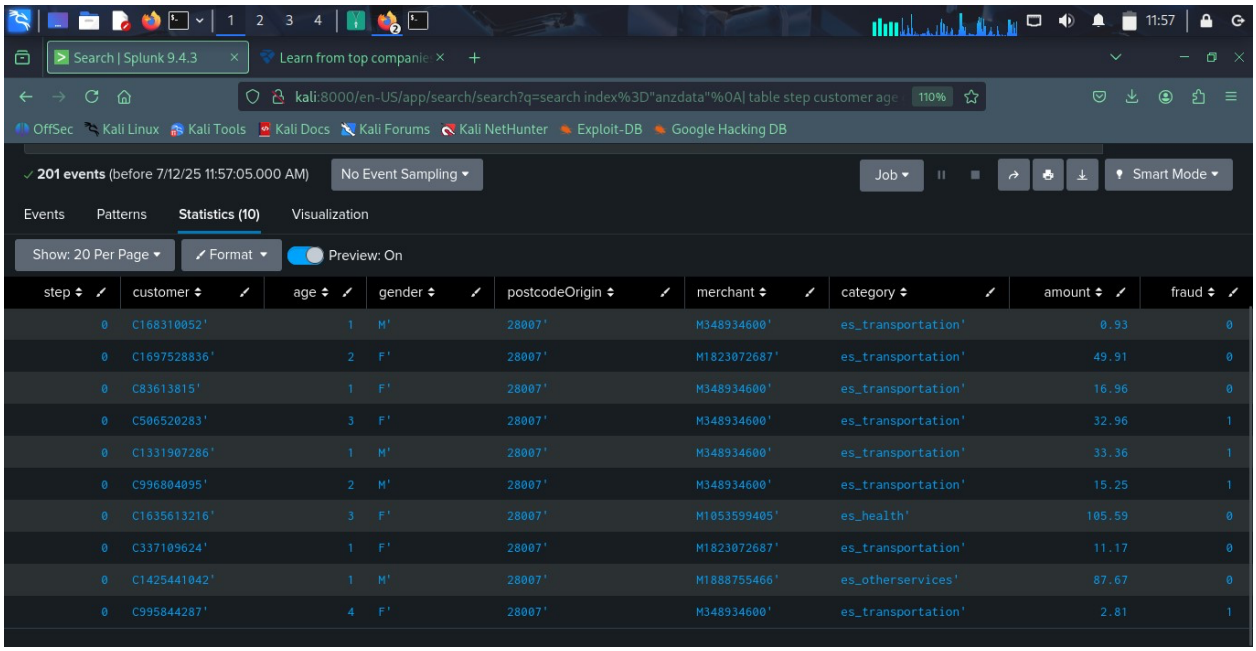


🧠 4. Data Normalization & Field Transformations

Field	Description	Transformation Applied
`step`	Activity month indicator (0–3)	Mapped to month name (May–Aug)
`age`	Customer age category (0.0–5.0)	Mapped to standard ranges (e.g., 26–35)
`gender`	F/M codes	Converted to "Female"/"Male"
`postcodeOrigin`	Origin postcode	Used for geo-analysis

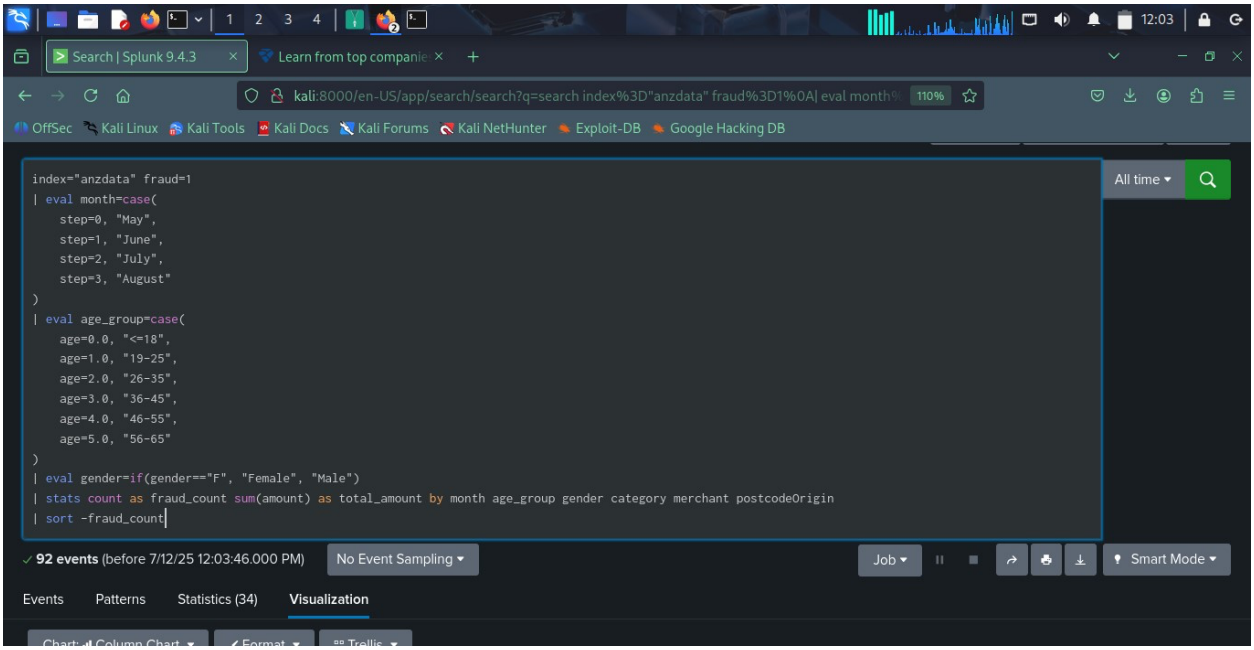
`merchant`	Unique merchant ID	Aggregated for fraud frequency

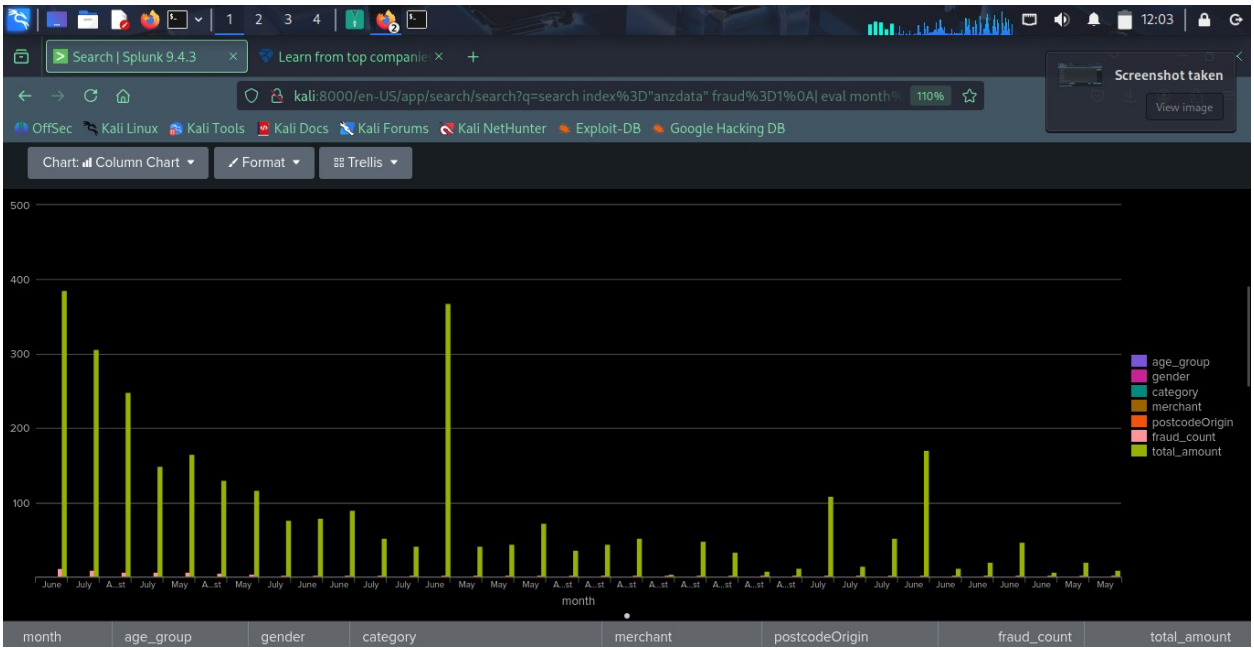
Field	Description	Transformation Applied
`category`	Purchase category (e.g., es_food)	Used to identify exploited verticals
`amount`	Transaction amount	Summed per group for fraud losses
`fraud`	Binary indicator (1 = fraud)	Filtered for fraud-only views

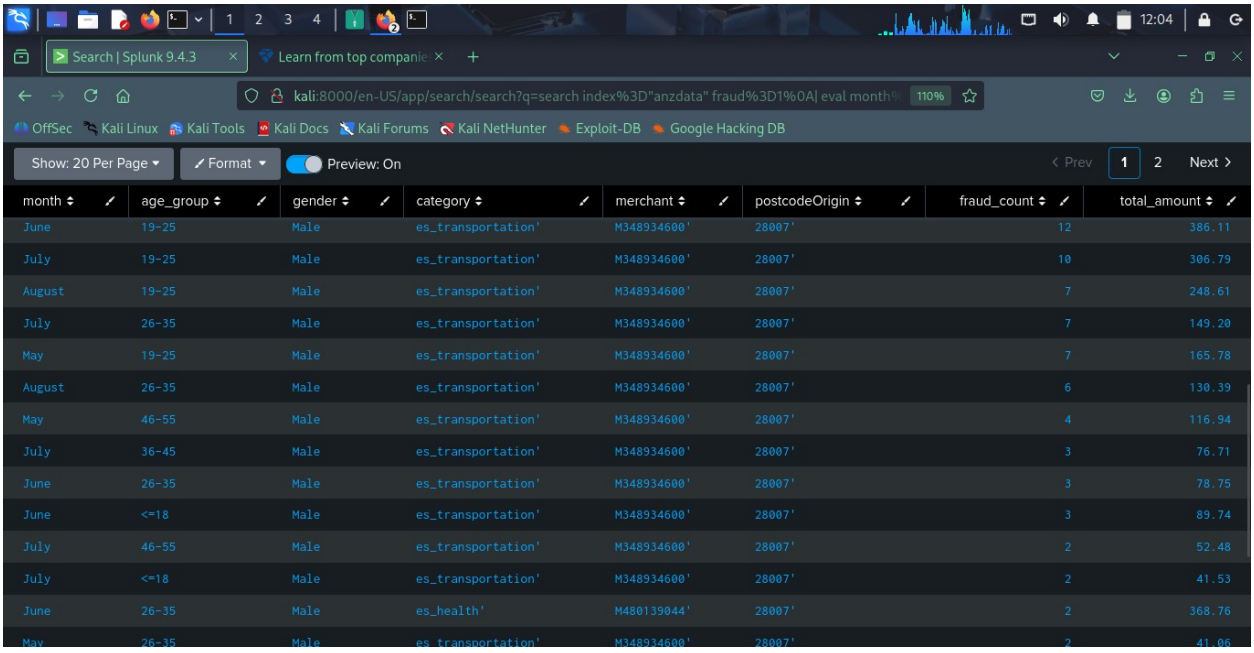


step	customer	age	gender	postcodeOrigin	merchant	category	amount	fraud
0	C168310052'	1	M'	28007'	M348934600'	es_transportation'	0.93	0
0	C1697528836'	2	F'	28007'	M1823072687'	es_transportation'	49.91	0
0	C83613815'	1	F'	28007'	M348934600'	es_transportation'	16.96	0
0	C506520283'	3	F'	28007'	M348934600'	es_transportation'	32.96	1
0	C1331907286'	1	M'	28007'	M348934600'	es_transportation'	33.36	1
0	C996804095'	2	M'	28007'	M348934600'	es_transportation'	15.25	1
0	C1635613216'	3	F'	28007'	M1053599405'	es_health'	105.59	0
0	C337109624'	1	F'	28007'	M1823072687'	es_transportation'	11.17	0
0	C1425441042'	1	M'	28007'	M1888755466'	es_otherservices'	87.67	0
0	C995844287'	4	F'	28007'	M348934600'	es_transportation'	2.81	1

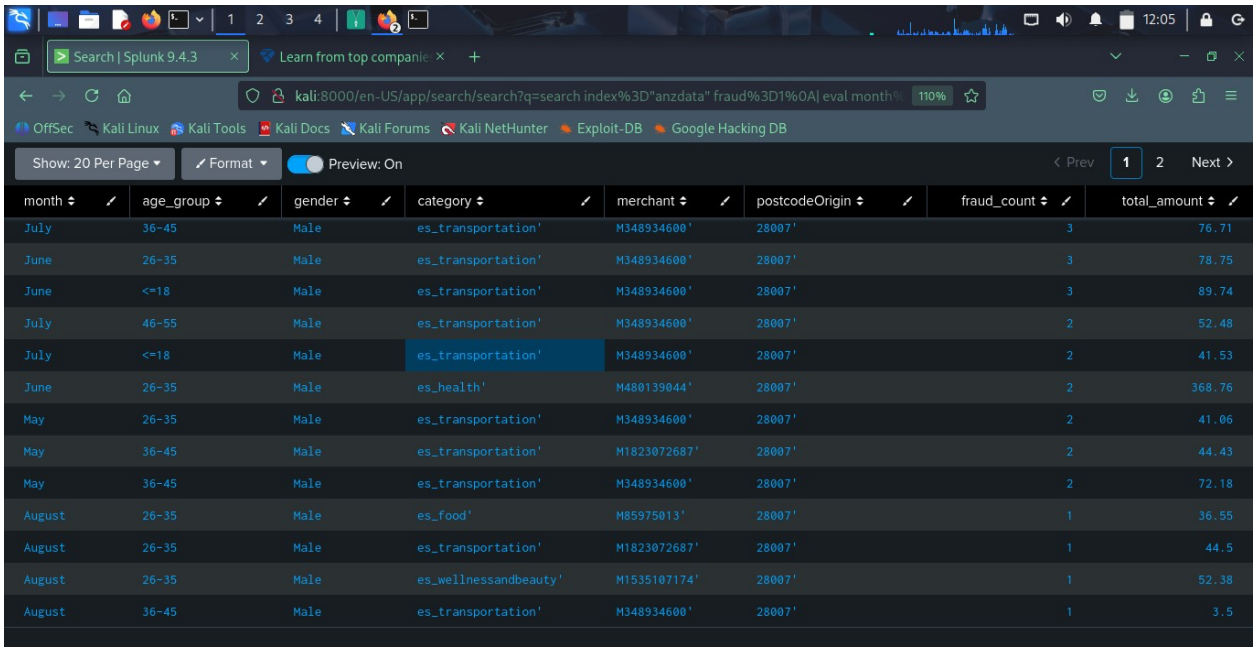

Search | Splunk 9.4.3
Learn from top companie
kali:8000/en-US/app/search/search?q=search index%3D"anzdata" fraud%3D1%0A| stats count b
110%
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
50
40
30
20
10
F'
M'
M'
F'
F'
F'
M'
gender
category
count
gender
category
count
F'
es_transportation'
49
M'
es_transportation'
35

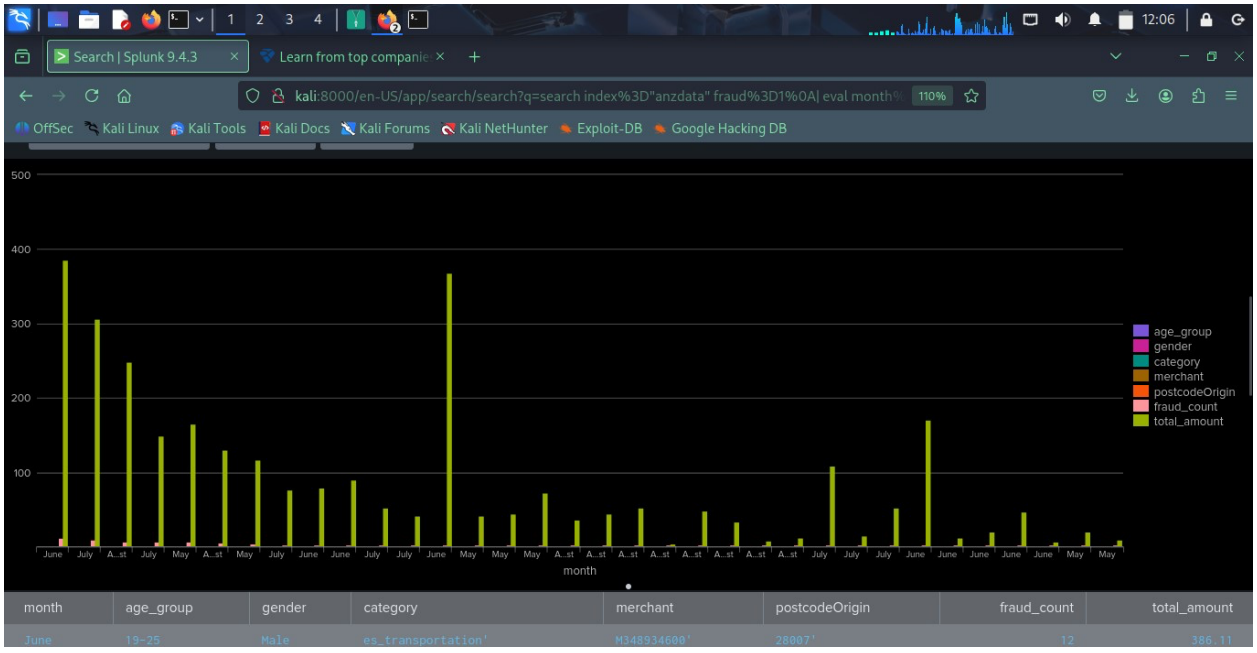

Search | Splunk 9.4.3
Learn from top companie
kali:8000/en-US/app/search/search?q=search index%3D"anzdata" fraud%3D1%0A| eval month%
110%
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
index="anzdata" fraud=1
| eval month=case(
step=0, "May",
step=1, "June",
step=2, "July",
step=3, "August"
)
| eval age_group=case(
age=0.0, "<=18",
age=1.0, "19-25",
age=2.0, "26-35",
age=3.0, "36-45",
age=4.0, "46-55",
age=5.0, "56-65"
)
| eval gender=if(gender=="F", "Female", "Male")
| stats count as fraud_count sum(amount) as total_amount by month age_group gender category merchant postcodeOrigin
| sort -fraud_count
All time
92 events (before 7/12/25 12:03:46.000 PM)
No Event Sampling
Job
Smart Mode
Events
Patterns
Statistics (34)
Visualization





month	age_group	gender	category	merchant	postcodeOrigin	fraud_count	total_amount
June	19-25	Male	es_transportation'	M348934600'	28007'	12	386.11
July	19-25	Male	es_transportation'	M348934600'	28007'	10	306.79
August	19-25	Male	es_transportation'	M348934600'	28007'	7	248.61
July	26-35	Male	es_transportation'	M348934600'	28007'	7	149.20
May	19-25	Male	es_transportation'	M348934600'	28007'	7	165.78
August	26-35	Male	es_transportation'	M348934600'	28007'	6	130.39
May	46-55	Male	es_transportation'	M348934600'	28007'	4	116.94
July	36-45	Male	es_transportation'	M348934600'	28007'	3	76.71
June	26-35	Male	es_transportation'	M348934600'	28007'	3	78.75
June	<=18	Male	es_transportation'	M348934600'	28007'	3	89.74
July	46-55	Male	es_transportation'	M348934600'	28007'	2	52.48
July	<=18	Male	es_transportation'	M348934600'	28007'	2	41.53
June	26-35	Male	es_health'	M480139044'	28007'	2	368.76
May	26-35	Male	es_transportation'	M348934600'	28007'	2	41.06





🔎 5. Core SPL Query for Unified Fraud View

```
spl
index="anzdata" fraud=1
| eval month=case(step=0,"May",step=1,"June",step=2,"July",step=3,"August")
| eval age_group=case(age=0.0,"<=18",age=1.0,"19-25",age=2.0,"26-
35",age=3.0,"36-45",age=4.0,"46-55",age=5.0,"56-65")
| eval gender=if(gender=="F","Female","Male")
| stats count as fraud_count sum(amount) as total_amount by month age_group
gender category merchant postcodeOrigin
```

```
| sort -fraud_count
```

📈 6. Key Findings

👤 Demographics

- **Highest fraud activity:** Males aged **19–25** and **26–35**
- **Peak activity months:** June and July for the 19–25 group

🧾 Exploited Categories

- **#1 Target:** `es_transportation`
- Other categories: `es_food`, `es_health`, `es_wellnessandbeauty`

🛒 Top Merchant Targets

- **Merchant M348934600** appeared consistently across age groups/months
- **Merchant M1823072687** showed repeat fraud at a lower volume

🌍 Geographic Origin

- Most fraudulent activity originated from **postcode 28007**

💸 Fraud Impact Snapshot

- For a single merchant/category/month:
  - Over **12 confirmed fraud cases** ○ Exceeding **$380 in losses** ○ Profile: **Male**, age **19–25**, month **June**

■7. Recommended Dashboard Panels

Panel Title	Visualization Type	Fields Used
Fraud Count by Age Group	Bar Chart	`age_group`, `fraud_count`
Fraud by Gender and Category	Column Chart	`gender`, `category`, `fraud_count`
Top Merchants by Fraud	Table / Bar Chart	`merchant`, `fraud_count`, `amount`
Monthly Fraud Trends	Line Chart	`month`, `fraud_count`
Geographic Fraud Heatmap (Opt.)	Map	`postcodeOrigin`, `fraud_count`

8. Recommendations & Next Steps

- 🔍 Expand dataset to include **non-fraud** records to compute **fraud ratios**
- 📣 Set **alerts** for sudden spikes by merchant or category
- 🧠 Use **Splunk ML Toolkit** for behavioral anomaly detection
- 🌐 Cluster analysis across **postcode** zones
- 📂 Automate ingestion using `inputs.conf` and monitored folders

📁

📁 9. Appendix: Field Mappings

```
yaml
step:   0
= May
1 = June
2 = July
3 = August
 age:   0.0 =
<= 18
  1.0 = 19-25
  2.0 = 26-35
  3.0 = 36-45
  4.0 = 46-55
  5.0 = 56-65
gender:   F =
Female
  M = Male
```

What This Shows

This project validates my **proficiency in Splunk SPL**, **data normalization**, and **SOC-ready dashboard design**. It also highlights my ability to troubleshoot ingestion issues, align datasets with business intelligence goals, and derive **real-time threat insights** from raw data.
