

# Introduction to Data Science – Spring 1404

Mehrnaz Hosseini 220701051

Helia Ranjbar 220701062

For the Data Ingestion section, after pulsing the data in the darooghe.transactions kafka topic, the validator check the values and add them to its proper topic, based on validation result.

## 1. Error\_log message and darooghe.error\_logs kafka topic

```
{"transaction_id": "6aae0cf2-d1b9-4597-9c51-a3aaa81b3531", "errors": ["ERR_AMOUNT"], "original_data": {"transaction_id": "6aae0cf2-d1b9-4597-9c51-a3aaa81b3531", "timestamp": "2025-05-15T09:18:35.759760", "customer_id": "cust_946", "merchant_id": "merch_46", "merchant_category": "transportation", "payment_method": "pos", "amount": 587171.0, "location": {"lat": 35.67438022274483, "lng": 51.357254364218434}, "device_info": {}, "status": "approved", "commission_type": "flat", "commission_amount": 11743.0, "vat_amount": 52845.0, "total_amount": 640016.0, "customer_type": "business", "risk_level": 1, "failure_reason": null}}
```

```
Invalid: 2bee5d13-afae-403a-b607-863cdd0408d7 (['ERR_AMOUNT'])
Invalid: d589a5c5-3efe-4b89-8e92-c97e71657e84 (['ERR_AMOUNT'])
Invalid: 39843221-cf46-447e-813d-4919167016f3 (['ERR_AMOUNT'])
Invalid: 0e021acb-72d3-4f98-903e-a10eb0411f4b (['ERR_AMOUNT'])
```

## 2. darooghe.valid\_transaction which contain transaction after their validation is approved

```
{"transaction_id": "296b080e-d53c-4883-8382-6780213b14a0", "timestamp": "2025-05-15T09:22:45.888482", "customer_id": "cust_825", "merchant_id": "merch_42", "merchant_category": "entertainment", "payment_method": "online", "amount": 303237.0, "location": {"lat": 35.77174290959668, "lng": 51.35195854285697}, "device_info": {"os": "iOS", "app_version": "3.1.0", "device_model": "iPhone 15"}, "status": "approved", "commission_type": "flat", "commission_amount": 6064.0, "vat_amount": 27291.0, "total_amount": 336592.0, "customer_type": "CIP", "risk_level": 2, "failure_reason": null}
```

```
Valid: f24e683d-550d-413f-bf0c-1b481f5cf9bf - Stored in valid_transactions topic
Valid: f241207a-3373-4166-9db3-41dd203348a3 - Stored in valid_transactions topic
Valid: 76b01cec-b2d2-43ff-a4fd-1e2fcb6a607c - Stored in valid_transactions topic
Valid: b208b098-7897-48c5-b331-a6a72c96d3a3 - Stored in valid_transactions topic
```

### 3. Commission Analysis collections

- Commission\_efficiency

```
_id: ObjectId('6825b6b032a5ae36345382c7')
merchant_category : "retail"
total_commissions : 30331591
avg_commission : 20990.720415224914
commission_ratio : 0.01999953906546547
transaction_count : 1445
```

- Commission\_profitability

```
_id: ObjectId('6825b6b132a5ae36345382ce')
merchant_category : "entertainment"
commission_type : "tiered"
avg_commission : 20101.365422396855
net_profit : 5115670.85
transaction_count : 509
```

- optimal\_commission\_types

```
_id: ObjectId('6825b6b332a5ae36345382df')
merchant_category : "entertainment"
optimal_commission_type : "progressive"
estimated_profit : 5132057.1
```

### 4. Transaction Pattern Analysis collections

- temporal\_patterns

```
_id: ObjectId('6825b9fe608f774388fdd31f')
day_of_week : 4
hour : 9
time_of_day : "morning"
merchant_category : "retail"
transaction_count : 12
total_volume : 17205950
avg_amount : 1433829.1666666667
```

- peak\_transaction\_times

```
_id: ObjectId('6825b9ff608f774388fdd39e')
day_of_week : 5
hour : 9
time_of_day : "morning"
merchant_category : "food_service"
transaction_count : 1691
total_volume : 1742195398
avg_amount : 1030275.2205795387
```

- customer\_segments

```
_id: ObjectId('6825ba00608f774388fdd3aa')
customer_id : "cust_478"
transaction_frequency : 9
total_spent : 8955871
avg_transaction_size : 995096.7777777778
days_since_last_purchase : 0
segment : "High Value"
```

- merchant\_comparison

```
_id: ObjectId('6825ba02608f774388fdd794')
merchant_category : "food_service"
total_transactions : 2301
total_volume : 2369248976
avg_transaction_size : 1029660.5719252499
approval_rate : 0.943502824858757
```

- time\_of\_day\_analysis

```
_id: ObjectId('6825ba03608f774388fdd79b')
time_of_day : "morning"
merchant_category : "retail"
transaction_count : 1705
avg_amount : 1043024.8263929619
```

- spending\_trends

```
_id: ObjectId('6825ba04608f774388fdd7ad')
week : Object
  start : 2025-05-15T00:00:00.000+00:00
  end : 2025-05-22T00:00:00.000+00:00
  merchant_category : "entertainment"
  weekly_transactions : 1775
  weekly_volume : 1809135112
  previous_week_volume : 425441200
  volume_change : 1383693912
  is_increasing : true
```

## 5. Data Storage Implementation

We decided to partition data **by date** because our primary focus was **time-based analysis** rather than customer segmentation. Given the high variability in **customer types**, using them as a partition key would lead to excessive fragmentation, making queries less efficient. **Date-based partitioning**, on the other hand, ensures structured and logical data storage, allowing us to analyze **trends, peaks, and seasonal patterns** effectively.

## 6. Historical Aggregated Data

- Aggregated\_transactions\_daily\_merchant\_summary

```
_id: ObjectId('6825c3ab47138113de2e3b12')
date : "2025-05-15"
merchant_id : "merch_21"
merchant_category : "entertainment"
transaction_count : 69
daily_volume : 67534587
avg_transaction_amount : 978762.1304347826
```

- Aggregated\_transactions\_weekly\_customer\_segments

```
_id: ObjectId('6825c3ab47138113de2e3d0c')
year : 2025
week_num : 20
customer_type : "individual"
unique_customers : 998
weekly_volume : 6286988181
week : "2025-20"
```

- Aggregated\_transactions\_monthly\_commission\_reports

```
_id: ObjectId('6825c3ac47138113de2e3d19')
month : "2025-05"
merchant_category : "entertainment"
total_commission : 72987730
avg_commission : 20251.867369589345
commission_ratio : 0.019999516297518234
```

- Aggregated\_transactions\_customer\_profiles

```
_id: ObjectId('6825c3ad47138113de2e3d28')
customer_id : "cust_1"
avg_transaction_amount : 827824.125
stddev_transaction_amount : 621026.9627766202
total_transactions : 8
max_transaction_amount : 1852031
min_transaction_amount : 179221
customer_type : "individual"
```

## 7. Fraud Detection

```
{"fraud_type":"AMOUNT_ANOMALY","fraud_description":"Customer cust_372 made transaction of 1958690.0 which is >1000% of their average (981,574.93)","fraud_timestamp":"2025-05-15T17:35:32.718+03:30","transactions":[{"transaction_id":"2c930cc5-a514-4263-bf05-0430d12218f6","timestamp":"2025-05-15T14:05:23.706+03:30","customer_id":"cust_372","merchant_id":"merch_50","merchant_category":"government","payment_method":"nfc","amount":1958690.0,"location":{"lat":35.73007484039816,"lng":51.309846992889824},"device_info":{"status":"approved","commission_type":"flat","commission_amount":39173.0,"vat_amount":176282.0,"total_amount":2174145.0,"customer_type":"business","risk_level":2,"latitude":35.73007484039816,"longitude":51.309846992889824,"avg_transaction_amount":981574.9333333333,"stddev_transaction_amount":620541.1268161116,"total_transactions":15,"is_anomaly":true,"fraud_type":"AMOUNT_ANOMALY","fraud_description":"Customer cust_372 made transaction of 1958690.0 which is >1000% of their average (981,574.93)","fraud_timestamp":"2025-05-15T17:35:32.718+03:30"}]}
```

## 8. darooghe.insights using checkpoints

```
{"window":{"start":"2025-05-15T12:48:20.000+03:30","end":"2025-05-15T12:49:20.000+03:30"},"merchant_category":"entertainment","transaction_count":40,"avg_amount":102072.975,"total_commission":816037.0}
{"window":{"start":"2025-05-15T12:48:20.000+03:30","end":"2025-05-15T12:49:20.000+03:30"},"merchant_category":"transportation","transaction_count":49,"avg_amount":1053137.6530612244,"total_commission":1032054.0}
{"window":{"start":"2025-05-15T12:48:20.000+03:30","end":"2025-05-15T12:49:20.000+03:30"},"merchant_category":"food_service","transaction_count":46,"avg_amount":979458.4565217391,"total_commission":901083.0}
```

## 9. Real-time commission analysis

- Darooghe.comission\_by\_type

```
{"window_start":"2025-05-15T12:47:40.000+03:30","window_end":"2025-05-15T12:48:40.000+03:30","commission_type":"flat","total_commission_by_type":1462884.0,"transaction_count":79}
{"window_start":"2025-05-15T12:48:20.000+03:30","window_end":"2025-05-15T12:49:20.000+03:30","commission_type":"flat","total_commission_by_type":1753121.0,"transaction_count":84}
{"window_start":"2025-05-15T12:47:40.000+03:30","window_end":"2025-05-15T12:48:40.000+03:30","commission_type":"tiered","total_commission_by_type":1317705.0,"transaction_count":66}
```

- Darooghe.comission\_ratio

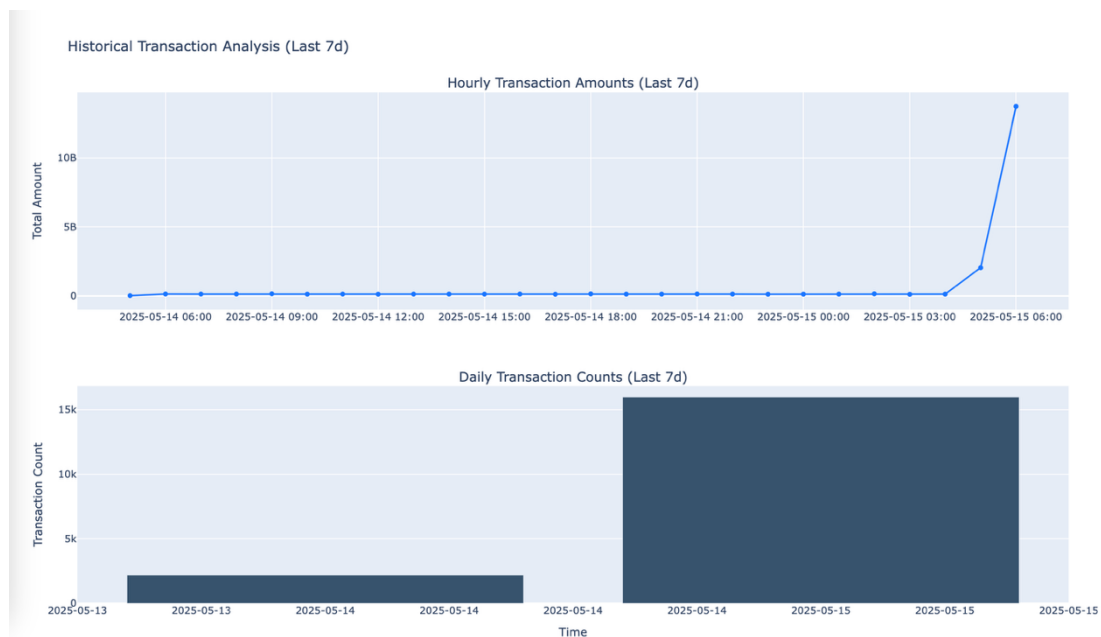
```
{"window_start":"2025-05-15T12:47:40.000+03:30","window_end":"2025-05-15T12:48:40.000+03:30","merchant_category":"entertainment","commission_ratio":0.019999485272709616,"total_commission":755332.0,"total_amount":3.7767572E7}
{"window_start":"2025-05-15T12:47:20.000+03:30","window_end":"2025-05-15T12:48:20.000+03:30","merchant_category":"food_service","commission_ratio":0.01999954780386499,"total_commission":647492.0,"total_amount":3.2375332E7}
{"window_start":"2025-05-15T12:48:00.000+03:30","window_end":"2025-05-15T12:49:00.000+03:30","merchant_category":"entertainment","commission_ratio":0.019999486033628274,"total_commission":933111.0,"total_amount":4.6656749E7}
```

- darooqhe.top\_commission\_merchants

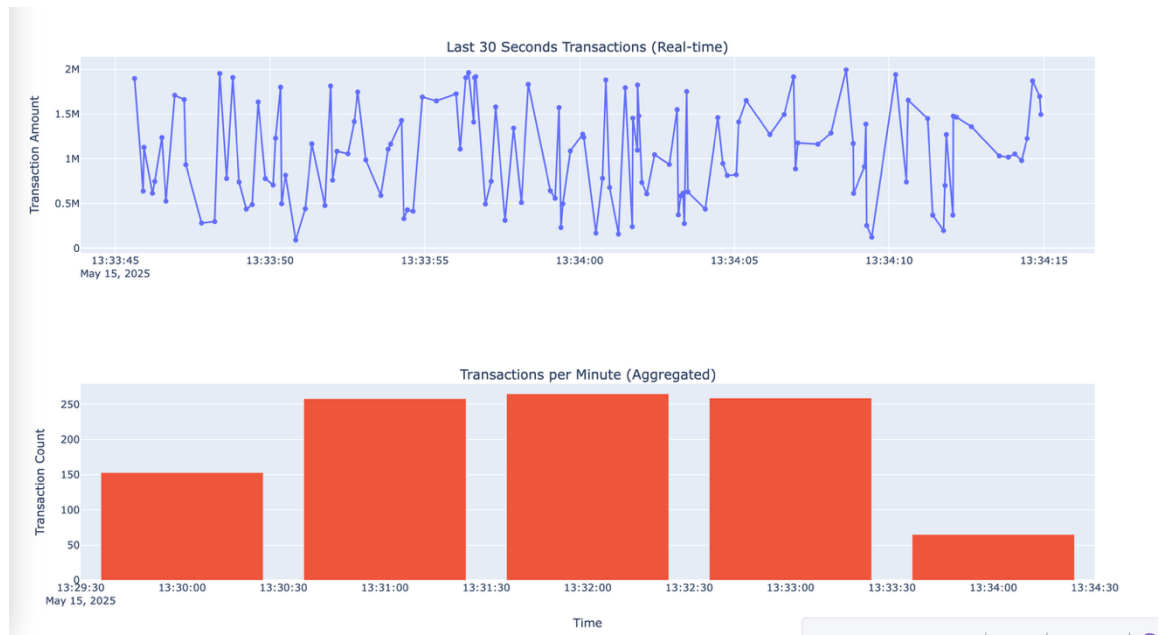
```
{"window_start":"2025-05-15T12:45:00.000+03:30","window_end":"2025-05-15T12:50:00.000+03:30","merchant_id":"merch_27","total_commission":7131.0,"transaction_count":1}
```

## 10. Visualization

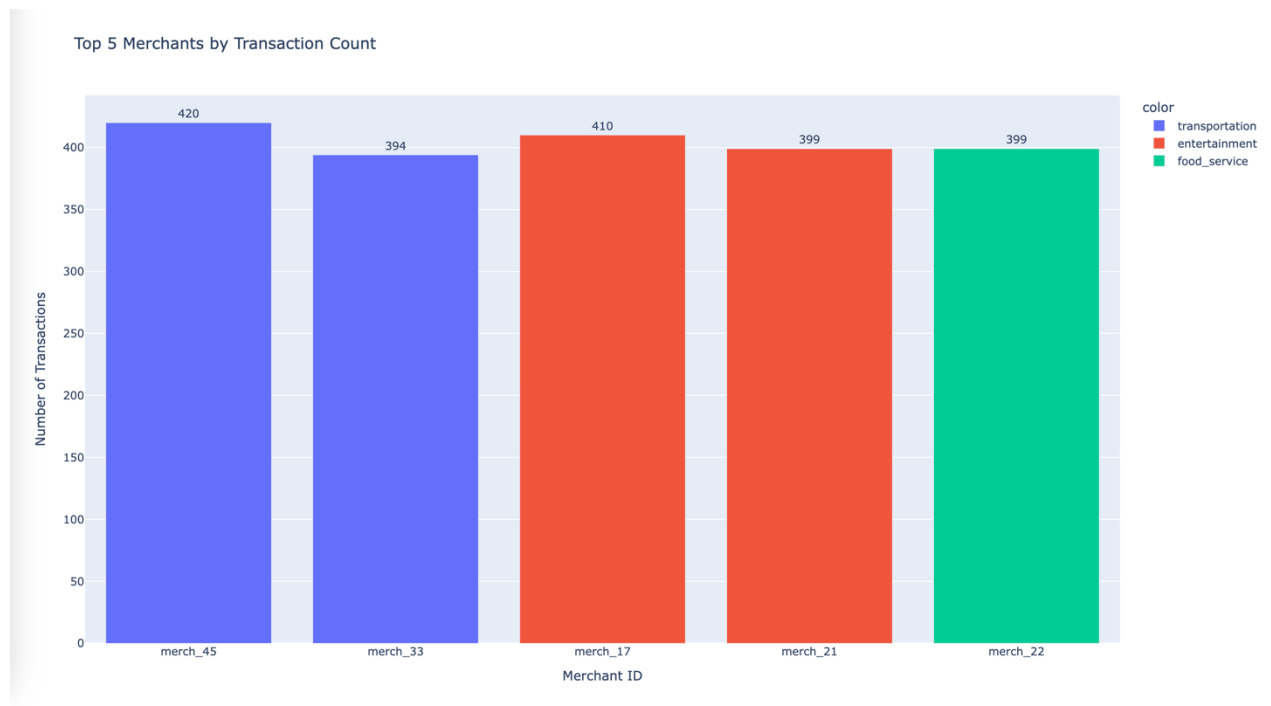
- Historical\_transaction\_volume



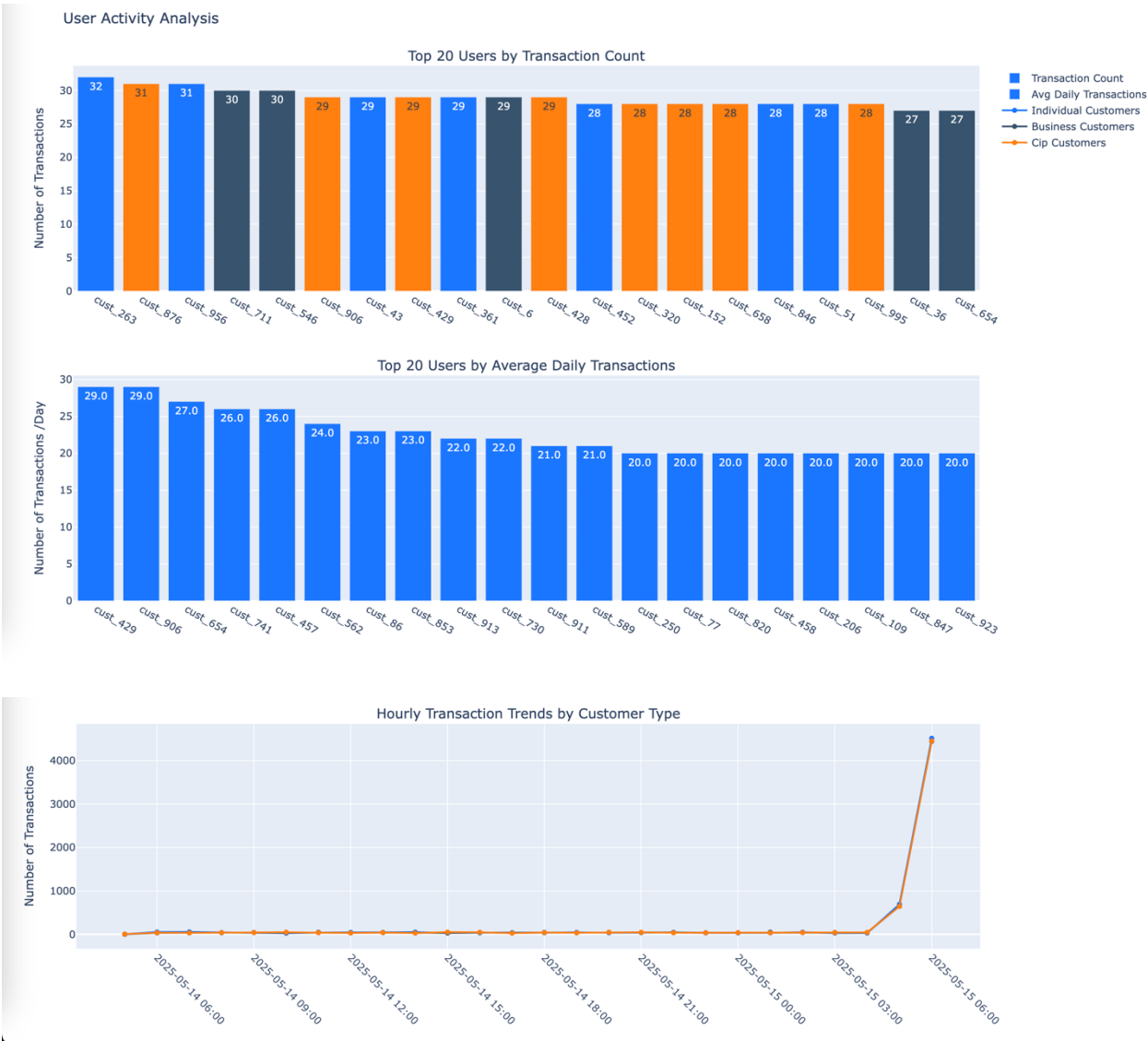
- Real-time volume



- Merchant\_analysis



- User\_activity\_analysis





## 11. Temporal Analysis

- merchant\_business\_hours\_analysis

```
_id: ObjectId('6825c0ee71c9e75e8a49e5e4')
amount : 511917
commission_amount : 10238
commission_type : "tiered"
customer_id : "cust_271"
customer_type : "CIP"
▼ device_info : Object
  app_version : "1.9.5"
  device_model : "Google Pixel 6"
  os : "Android"
  failure_reason : "system_error"
▼ location : Object
  lat : 35.682629534888235
  lng : 51.3632505082372
merchant_category : "food_service"
merchant_id : "merch_40"
partition_date : "20250514"
payment_method : "mobile"
risk_level : 3
status : "declined"
timestamp : 2025-05-14T16:21:55.739+00:00
total_amount : 568227
transaction_id : "752615e8-337b-4863-830d-7a3c8a50cbdf"
vat_amount : 46072
latitude : 35.682629534888235
longitude : 51.3632505082372
outside_business_hours : true
```

- merchant\_category\_activity

```
_id: ObjectId('68152b73cccc984d092d0d2a')
merchant_category : "entertainment"
time_segment : "Late Night (12AM-6AM)"
start_hour : 0
transaction_count : 155
total_amount : 160533902
avg_amount : 1035702.593548387
```

- merchant\_transaction\_spikes

```
_id: ObjectId('68152b74cccc984d092d0d44')
merchant_id : "merch_9"
merchant_category : "government"
spike_count : 2
peak_volume : 65
avg_normal_volume : 56.5
last_spike_time : 2025-04-29T12:30:00.000+00:00
```

