



Fraud Alert Exploratory Data Analysis

From Raw Alerts to Actionable Insights

Manmeet Singh • September 2025



Project Overview

Objective

Explore fraud alert data to uncover behavioral and velocity patterns that distinguish fraudulent from legitimate transactions

Dataset

Synthetic alert-level data containing transaction details, behavioral metrics, and velocity features across multiple time windows

Goal

Build a reproducible, explainable analytical pipeline optimized for future machine learning integration and stakeholder communication

Data Challenges

Primary Issues Identified

- Missing values and inconsistent data formats requiring extensive preprocessing
- No clear separation between fraud and non-fraud alerts in raw dataset
- Velocity features contained significant noise and lacked interpretability
- Behavioral metrics missing critical context for analysis

transaction_id	user_id	event_time	amount	currency	merchant	merchant_device_ip	addr	country	lat	lon	channel	payment_type	velocity	velocity	geo_dist	behavior	behavior	is_fraud	fraud_type
8a14553	2378	2025-01-08T08:01:00Z	88.81	INR	95293	electronics	dev-237	62.8.18..SM	17.8	103.8	pos	wallet	9	9	47.13	4.615	1.484	1	bot_burst
25d683	146	2025-01-08T08:08:00Z	208	INR	88613	electronics	dev-146	161.178.1.80	23.83	52.37	web	netbanki	18	18	28.08	6.701	2.127	1	bot_burst
1cd34a	334	2025-01-08T08:14:00Z	467.6	INR	43058	electronics	dev-334	12.124.1.41..GN	20.5	78.82	pos	netbanki	4	4	26.83	5.347	2.164	0	
d77de9f	1263	2025-01-08T08:20:00Z	563.4	INR	44711	travel	fds3f33	193.5.30.1..TO	23.67	53.53	pp	api	26	26	20.47	7.826	2.212	0	
ec733df	4537	2025-01-08T08:25:00Z	255.6	INR	38014	crypto	dev-4537	30.9.32.1..MH	23.57	54.13	web	netbanki	23	23	63	5.768	2.604	1	bot_burst
c5f1420	2712	2025-01-08T08:34:00Z	183.4	INR	95344	food	dev-2712	13.43.2..TJ	20.36	79.05	pos	netbanki	21	21	33.78	8.061	2.234	1	bot_burst
8b51126	1228	2025-01-08T08:40:00Z	54.0	INR	84338	utilities	dev-122	92.154.5..PW	1.46	103.8	pp	card	19	19	85.43	6.552	1.368	1	bot_burst
6e575b6	10	2025-01-08T08:45:00Z	588.3	INR	43631	electronics	dev-10	87.165.6..BT	31.42	-35.9	pp	wallet	34	34	37.2	5.191	2.024	1	bot_burst
5bf233f	334	2025-01-08T08:50:00Z	350.5	INR	52683	crypto	dev-3294	117.172.1..CG	20.52	78.57	pp	card	5	5	25.34	3.935	2.113	1	bot_burst
423db2	1853	2025-01-08T08:52:00Z	82.0	INR	95673	gaming	dev-185	220.234..LY	54.8	-3.485	web	wallet	36	36	13.43	4.324	1.84	1	bot_burst
c157e0c	3037	2025-01-08T08:57:00Z	283.3	INR	60611	fashion	dev-186	177.187.5..VC	37.3	-35.24	web	api	1	1	14.45	6.31	2.204	0	
75d88f7	4167	2025-01-08T08:58:00Z	100.8	INR	34572	gaming	dev-416	86.3.87..TM	55.65	-3.295	web	api	31	31	3.034	4.228	1.714	0	
8e6a316	4853	2025-01-08T08:59:00Z	188	INR	70617	retail	dev-9533	112.131.2..CI	36.63	-36.28	pp	api	23	23	2.792	5.548	2.691	1	bot_burst
r6225dt	4554	2025-01-08T09:00:00Z	214.4	INR	44380	crypto	dev-455	149.21..GE	23.24	53.48	web	card	23	23	26.78	5.495	1.686	1	bot_burst
b24804	2304	2025-01-08T09:04:00Z	79.8	INR	1033	gaming	dev-293	193.24.1..LK	1.03	104.1	pp	netbanki	12	12	42.18	6.39	0.826	1	bot_burst
34d13d	926	2025-01-08T09:07:00Z	117	INR	63932	travel	dev-3882	170.203..SY	37.38	-36.58	pos	card	11	11	27.33	6.579	2.203	1	bot_burst
c9524c	3653	2025-01-08T09:14:00Z	142.4	INR	81463	travel	dev-365	160.183..AZ	20.54	78.47	pp	wallet	36	36	9.401	3.515	1.943	1	bot_burst
3979715	4628	2025-01-08T09:15:00Z	463.3	INR	70529	travel	dev-521	154.8.0..TO	55.35	-3.09	pp	card	1	1	28.92	5.823	2.312	0	
cdb2e5	778	2025-01-08T09:16:00Z	31.3	INR	44534	travel	dev-3181	45.153.2..SM	23.34	53.45	web	api	27	27	8.339	4.493	3.086	0	
5593e51	2045	2025-01-08T09:17:00Z	68.6	INR	55530	gaming	dev-2392	57.43.16..EG	55.03	-3.574	pp	card	22	22	29.64	3.347	2.535	1	bot_burst
c6d648	2813	2025-01-08T09:18:00Z	867.4	INR	10873	travel	dev-281	102.195..MX	23.47	54.23	pp	api	13	13	40.82	5.77	1.9	1	bot_burst
bf2fc0f	819	2025-01-08T09:19:00Z	505.6	INR	71111	utilities	dev-819	53.130.0..BD	23.26	53.7	pp	card	34	34	6.435	6.026	0.394	1	bot_burst
05d60c	4670	2025-01-08T09:21:00Z	34.8	INR	70482	food	dev-467	57.223..LY	37.01	-95.78	pos	wallet	10	10	34.03	5.267	1.327	1	bot_burst
4200bf	375	2025-01-08T09:23:00Z	238.5	INR	37418	fashion	dev-375	10.21.14..PK	23.25	53.64	pos	card	31	31	8.156	6.416	2.51	1	bot_burst
84f4cc	1922	2025-01-08T09:25:00Z	52.01	INR	60032	travel	dev-427b	3.138.17..BO	31.23	-35.39	pp	api	12	12	75.33	4.201	2.763	1	bot_burst
02214fc	4259	2025-01-08T09:26:00Z	516	INR	76060	food	dev-425	81.90.17..IT	1.01	103.9	pos	wallet	18	18	5.939	5.497	2.645	1	bot_burst
ba66f6d	391	2025-01-08T09:27:00Z	158.5	INR	88570	travel	cb9618	41.39.43..BO	36.36	53.43	web	netbanki	26	26	29.2	6.197	2.466	0	
36d161	57	2025-01-08T09:30:00Z	403.4	INR	93771	electron	dev-1230	55.49.2..CI	20.29	79.04	pp	api	38	38	24.66	4.771	1.405	1	bot_burst
514bcf	2040	2025-01-08T09:31:00Z	518.	INR	82094	travel	dev-60	156.127..HR	23.85	53.74	pp	wallet	28	28	25.03	5.448	1.53	1	bot_burst
434ea7f	2981	2025-01-08T09:33:00Z	33.75	INR	47483	food	dev-296	220.146..UG	0.419	103.3	pp	wallet	9	9	19.93	4.889	2.821	1	bot_burst
e41fc48	541	2025-01-08T09:34:00Z	1126	INR	53134	travel	dev-541	92.221..JM	55.58	-3.443	pp	card	1	1	13.53	3.414	1.833	0	
b3f8956	1586	2025-01-08T09:37:00Z	96.3	INR	88611	gaming	dev-1581	36.171..H..W	1.52	103.6	pos	wallet	6	6	18.52	4.586	1.62	1	bot_burst
e656ebc	3665	2025-01-08T09:41:00Z	47.51	INR	63930	retail	dev-365	175.42.2..GA	20.24	73.73	pos	api	4	4	29.3	6.211	2.16	0	
204f19f	3806	2025-01-08T09:42:00Z	562.7	INR	47198	utilities	dev-38	42.142.1..IS	20.63	78.96	web	api	28	28	41.62	3.397	2.305	1	bot_burst
43919f7	28	2025-01-08T09:43:00Z	70.9	INR	79878	food	bd054c	92.32.8..TJ	20.06	79.48	web	card	57	57	16.6	3.11	2.007	0	
da29b0	3390	2025-01-08T09:44:00Z	633.6	INR	57612	travel	dev-335	65.8.53..CI	55.31	-3.437	web	api	31	31	75.37	5.191	1.373	1	bot_burst
9e052d1	170	2025-01-08T09:45:00Z	162.6	INR	63563	gaming	dev-103	8.201.26..EG	37.23	-35.33	pp	card	7	7	21.25	4.216	1.669	1	bot_burst
2526cf7	2811	2025-01-08T09:46:00Z	73.24	INR	55916	fashion	dev-304	59.97.85..SY	20.83	83.85	web	netbanki	29	29	18.32	2.095	2.881	1	bot_burst
c5898ca	1130	2025-01-08T09:47:00Z	616.6	INR	84602	travel	dev-113	85.62.23..GE	0.315	104.4	pp	card	20	20	104.2	4.553	2.119	0	
c23f0f5	3527	2025-01-08T09:48:00Z	725.3	INR	42358	crypto	0b79d0	66.64.62..IR	20.22	73.25	web	netbanki	0	0	0	4.936	2.662	0	
37175	2025-01-08T09:49:00Z	198.2	INR	30554	food	121f14c	140.142.44..AM	23.29	53.58	pp	api	10	10	19.52	5.61	2.169	1	bot_burst	
d32cd6b	3135	2025-01-08T09:50:00Z	330.1	INR	64686	electronics	dev-213	11.176.2..AM	1.237	103.8	web	netbanki	22	22	4.772	5.83	2.603	1	bot_burst
359432f	2068	2025-01-08T09:51:00Z	214.3	INR	27435	travel	dev-204	66.214.1..SG	56.17	-3.22	pp	api	32	32	25.14	5.688	2.637	1	bot_burst
52159f1	4516	2025-01-08T09:52:00Z	197.1	INR	21692	fashion	dev-59	47.116.14..IN	55.16	-3.11	pp	api	17	17	74.17	5.674	2.220	1	bot_burst
s21566	144	2025-01-08T09:53:00Z	160.5	INR	46081	fuel	dev-445	124.10..TT	31.52	-35.57	web	netbanki	3	3	15.53	3.922	2.022	0	
s21567	2010	2025-01-08T09:54:00Z	254.5	INR	63521	food	dev-281	208.8.1..UY	22.06	54.3	pp	api	11	11	31.23	6.79	1.541	1	bot_burst
070047	2630	2025-01-08T09:55:00Z	41.39	INR	50531	food	dev-265	65.65.49..AU	21.05	70.37	pp	api	24	24	24.51	6.742	2.069	1	account_takeover
s2044dc	1615	2025-01-08T09:56:00Z	103	INR	73662	crypto	dev-101	81.115.41..SG	0.115	-36.11	pos	api	9	9	14.31	5.748	2.388	1	bot_burst
s2045dc	1006	2025-01-08T09:57:00Z	169.3	INR	42004	crypto	dev-100	124.54..PK	0.919	100.8	web	wallet	5	5	32.79	5.931	2.564	0	
26409c	2029	2025-01-08T09:58:00Z	561.4	INR	73016	gadget	dev-203	138.23..BG	1.16	104	pp	card	2	2	21.44	4.006	2.588	0	
190124f	1965	2025-01-08T09:59:00Z	196.7	INR	12124	gadget	dev-195	111.9.8..TD	31.23	-35.85	pp	api	45	45	35.95	5.963	2.354	1	bot_burst
91d3dc	3603	2025-01-09T00:01:00Z	144	INR	34446	fuel	dev-244	222.241..LC	55.41	-3.21	pp	card	9	9	8.119	4.438	1.509	1	bot_burst
s2045dc	2420	2025-01-09T00:02:00Z	604.4	INR	43456	travel	dev-245	134.15..SA	54.68	-3.616	pp	card	14	14	29.21	3.341	2.243	1	bot_burst
9541sf	773	2025-01-09T00:03:00Z	156.1	INR	18303														

Feature Engineering Strategy

01

Temporal Features

Created time-based indicators including hour of day, day of week, and weekend flags to capture temporal fraud patterns

02

Behavioral Metrics

Processed typing speed, navigation patterns, and geographic distance calculations to identify anomalous user behavior

03

Velocity Calculations

Engineered rolling window metrics for 1-hour, 24-hour periods, and 30-day unique device tracking

04

Anomaly Detection

Developed amount_over_user_avg metric to flag transactions significantly above user's historical spending patterns

```
# Step 6: Velocity and amount anomaly
print("\n▣ Computing transaction velocity and amount anomalies...")
df["tx_count_1d"] = df.groupby("user_id")["event_time"].transform(lambda x: x.expanding(1).sum())
df["tx_count_7d"] = df.groupby("user_id")["event_time"].transform(lambda x: x.expanding(7).sum())

rolling_avg = (
    df.set_index("event_time")
    .groupby("user_id")["amount"]
    .rolling("30d").mean()
    .rename("avg_amount_30d")
    .reset_index()
)
df = pd.merge(df, rolling_avg, on=["user_id", "event_time"], how="left")
df["avg_amount_30d"] = pd.to_numeric(df["avg_amount_30d"], errors="coerce").fillna(0)

df["amount_over_user_avg"] = df["amount"] / (df["avg_amount_30d"] + 1e-6)
df["amount_over_user_avg_log1p"] = np.log1p(df["amount_over_user_avg"])
df["is_amount_5x_user_avg"] = (df["amount_over_user_avg"] > 5).astype(int)
print("▣ Amount anomaly features computed.")

# Step 7: Unique devices in past 30 days
print("\n▣ Computing unique_devices_30d...")
df["unique_devices_30d"] = 0
user_groups = df.groupby("user_id")
for i, (user_id, group) in enumerate(user_groups):
    group = group.sort_values("event_time")
    result = []
    for j, row in group.iterrows():
        window_start = row["event_time"] - pd.Timedelta(days=30)
        window = group[(group["event_time"] >= window_start) & (group["event_time"] <= row["event_time"])]
        result.append(window["device_id"].nunique())
    df.loc[group.index, "unique_devices_30d"] = result
    if (i + 1) % 100 == 0 or i == len(user_groups) - 1:
        print(f"▣ Processed {i + 1} users...")
df["unique_devices_30d"] = pd.to_numeric(df["unique_devices_30d"], errors="coerce")
print("▣ Finished computing unique_devices_30d.")

# Step 8: Velocity features
print("\n▣ Computing velocity features (1h and 24h)...")
velocity_1h = (
    df.set_index("event_time")
    .groupby("user_id")["user_id"]
    .rolling("1h").count()
    .rename("velocity_1h")
    .reset_index()
)
velocity_24h = (
    df.set_index("event_time")
    .groupby("user_id")["user_id"]
    .rolling("1d").count()
    .rename("velocity_24h")
    .reset_index()
)

for vdf, name in [(velocity_1h, "velocity_1h"), (velocity_24h, "velocity_24h")]:
    dupes = vdf.duplicated(subset=["user_id", "event_time"], keep=False)
    if dupes.any():
        print(f"▲ Found {dupes.sum()} duplicate {name} rows - aggregating")
        vdf = vdf.groupby(["user_id", "event_time"], as_index=False).mean()
        if name not in vdf.columns:
            last_col = vdf.columns[-1]
            print(f"▲ Renaming column '{last_col}' to '{name}'")
            vdf.rename(columns={last_col: name}, inplace=True)

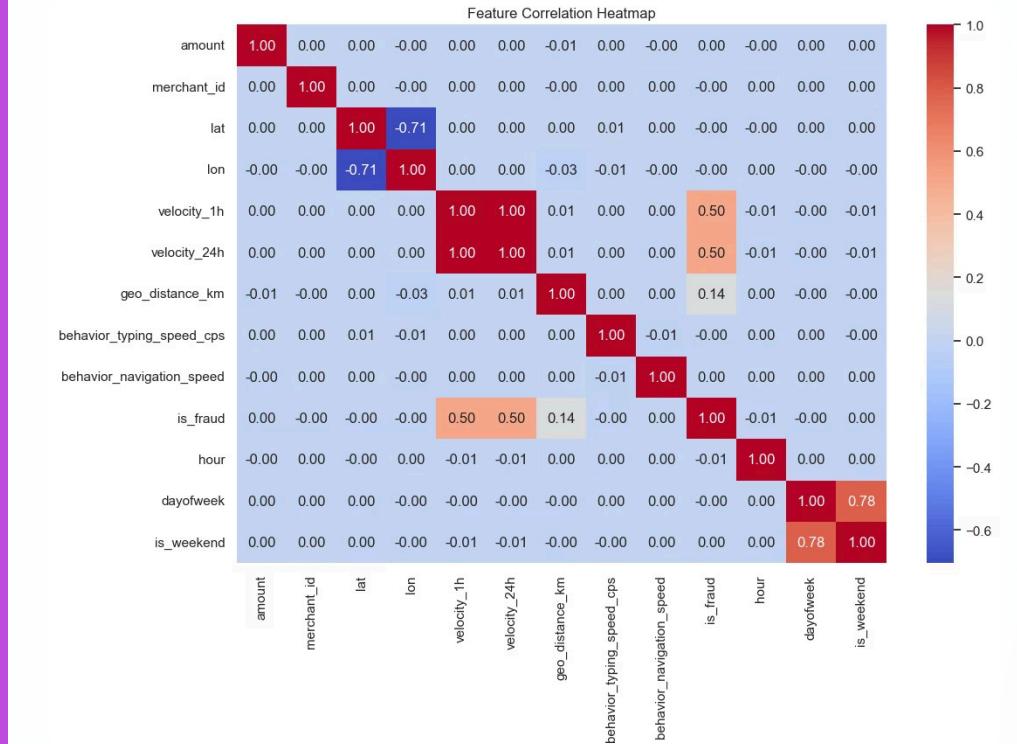
    df = pd.merge(df, vdf, on=["user_id", "event_time"], how="left")
    if name in df.columns:
        df[name] = pd.to_numeric(df[name], errors="coerce").fillna(0).astype(int)
        print(f"▣ {name} merged and cleaned.")
    else:
        print(f"✗ Merge failed: {name} not found")
        df[name] = 0
```

Exploratory Analysis Results

Comprehensive visualizations revealed clear patterns distinguishing fraudulent from legitimate transaction alerts

Analysis Components

- Distribution comparisons between fraud and legitimate alerts
- Time-of-day and weekend fraud pattern identification
- Velocity feature separation analysis
- Correlation mapping for all numeric features



Key Insights Discovered

Flagging high-risk transactions...
Alerts flagged using threshold 0.85

Data Quality Summary:
Total records: 200080
Missing values: 0 across 0 columns
Duplicate transactions: 80
Unique users: 5000
Unique devices: 64914
Fraud rate: 0.5293

Feature Distribution Overview:
amount min=3.60 max=15246.85 mean=296.54
typing_speed min=0.00 max=0.00 mean=0.00
nav_speed min=0.00 max=0.00 mean=0.00
geo_distance_km min=0.00 max=15463.43 mean=385.43
fraud_risk_score_weighted min=4.00 max=8.00 mean=4.62

Evaluation Metrics:

Precision & Ranking Metrics:

Metric Interpretation:
PR-AUC: Measures how well the model ranks frauds across all thresholds. Closer to 1 is better.
PR-AUC: 0.5718

ROC-AUC: Measures overall discrimination ability. 0.5 is random guessing.

ROC-AUC: 0.5274

Precision@1%: Of the top 1% scored transactions, this is the actual fraud rate.

Precision@1%: 0.8165

Top-k threshold: Minimum score required to be in the top 1% – useful for alerting.

Top-k threshold: 0.9925

Top 5 flagged alerts:

transaction_id	user_id	amount	merchant_category	model_score	fraud_risk_score_weighted
7ae6a220-01f1-4963-902f-8d5ddc86a186	513	88.75	fuel	0.997880	6
e9608552-f70e-4afc-85f1-03375f831e19	4097	102.00	fuel	0.997700	6
69c8a44d-97c1-456e-84c8-4a856d63b240	894	114.59	food	0.997588	6
cbf4d1-679f-41b0-a0e3-5b7ba9a1fd25	4883	152.74	food	0.997548	6
8cea7cb8-e432-4600-a914-9a0ecbd0d646	1181	306.99	utilities	0.997513	6

Saving flagged alerts to alerts.csv...
Saved all flagged alerts to alerts.csv

Total alerts flagged: 88091
Alert fraud rate: 0.5516

1

Fraud alerts exhibit measurably different velocity and behavioral patterns compared to legitimate transactions, creating clear separation opportunities

2

Late-night and weekend transactions show significantly higher fraud probability, indicating time-based risk factors

3

Velocity_1h and geo_distance_km emerged as the most powerful indicators for fraud detection models

4

Feature engineering dramatically improved data clarity and model-readiness, enabling more accurate fraud detection

Exported Output & Deliverables

1

Enhanced Dataset

All enriched alerts exported to alerts.csv with complete feature engineering applied for immediate ML model consumption

2

Reproducible Architecture

Modular Jupyter notebook structure designed for easy replication and modification across different datasets

3

Visual Reports

Comprehensive reports folder containing all analytical charts with detailed descriptions for stakeholder communication

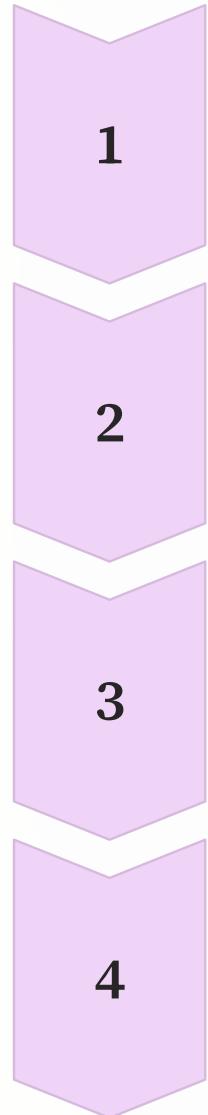
4

Documentation

Complete architecture documentation prepared for seamless future machine learning model integration

```
## 📁 Folder Structure
fraud-alert-analytics/
├── reports/
│   └── *.png, *.txt
├── data/
│   └── alerts.csv
│       └── synthetic_transaction.parquet
├── docs/
│   └── architecture.md
├── logs/
│   └── model_metrics.txt
├── artifacts/
│   └── run_behavioural_fraud_pipeline.py
└── exploratory_analysis.ipynb
    ├── README.MD
    ├── requirements.txt
    ├── scripts
    └── xgb_behavioral_pipeline.joblib
```

Strategic Next Steps



Model Development

Build baseline ML models using logistic regression and random forest algorithms to establish performance benchmarks

Risk Scoring System

Implement risk tiering and weighted scoring logic to prioritize high-risk alerts for investigation teams

Automation Pipeline

Develop automated report generation and GitHub packaging system for continuous integration workflows

Stakeholder Engagement

Share comprehensive results with business stakeholders through interactive visual reports and dashboards

Machine Learning Model

Lecture lesson plan for the Machine Learning Model. This is a difficult learning exercise, but it will provide you with a solid foundation for your machine learning journey.



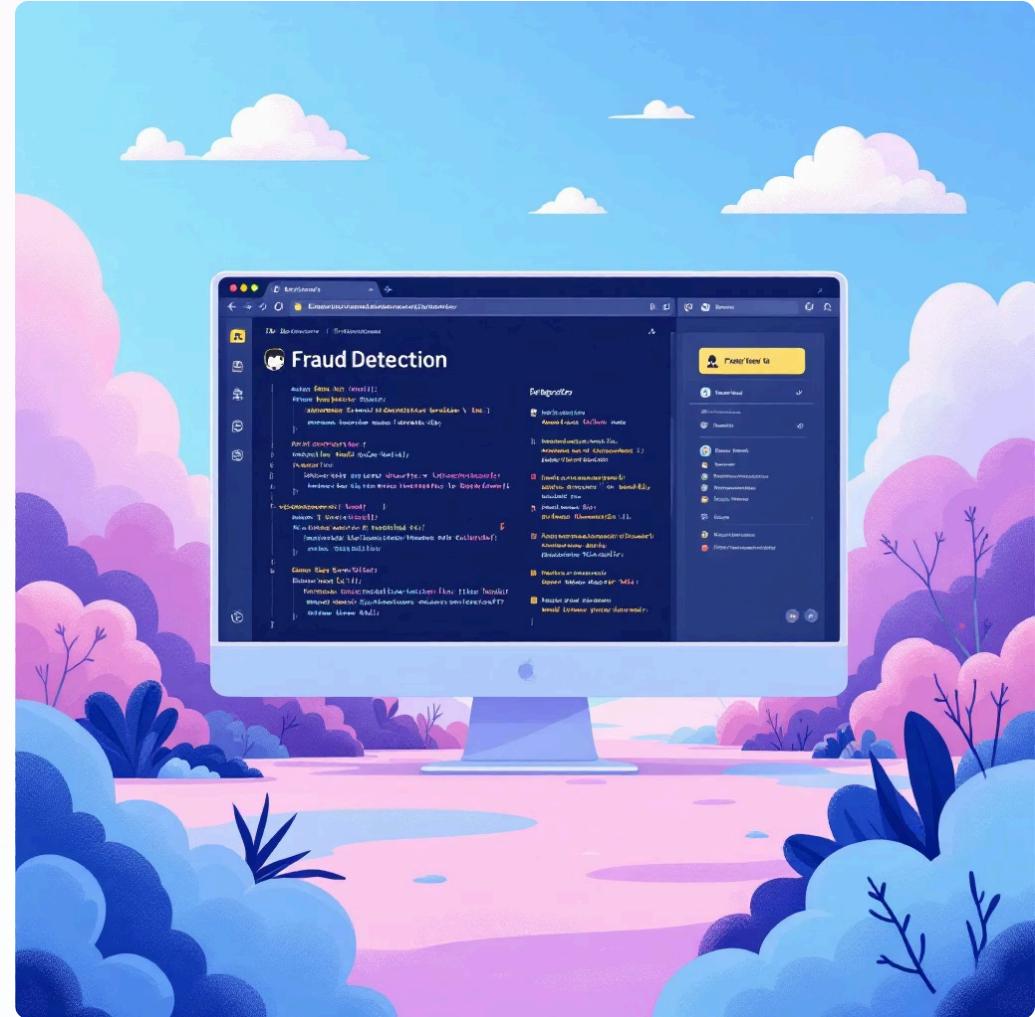
GitHub Repository Access

Complete Project Available

Access the full analytical pipeline, architecture documentation, and outputs through our comprehensive GitHub repository.

[View Repository](#)

Repository includes complete pipeline code, detailed architecture notes, sample outputs, and contribution guidelines. Open for feedback, forks, and collaborative improvements.





Thank You & Collaboration Opportunities

Velocity Metrics

How do you effectively handle velocity calculations across multiple rolling time windows in production environments?

GitHub Best Practices

What are your recommended strategies for packaging and presenting exploratory data science projects on GitHub?

Feature Scaling

Which techniques work best for scaling behavioral features while maintaining interpretability in fraud detection?

Let's Connect & Collaborate!