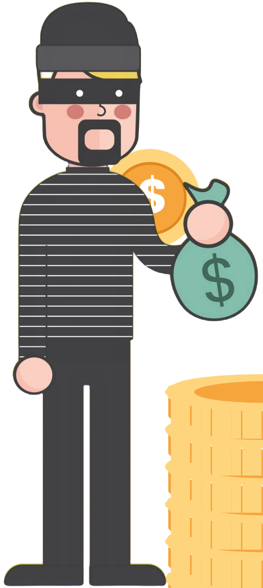
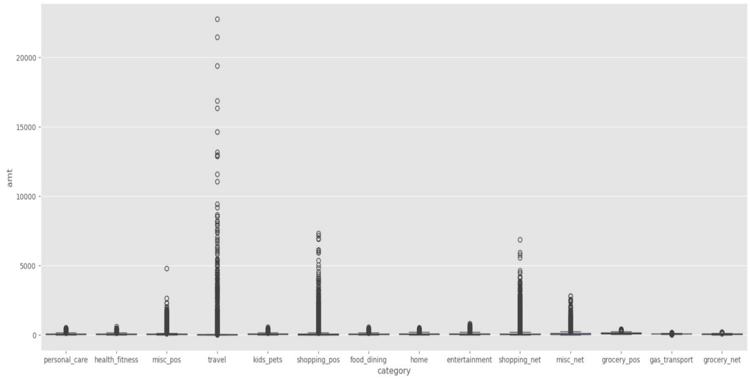
Credit Card Fraud Detection

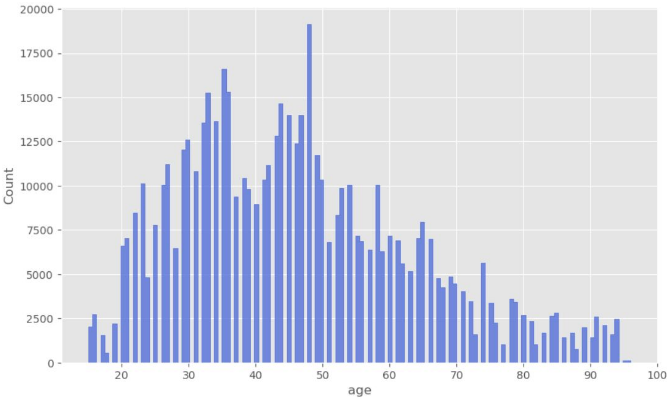
What patterns or behaviors differentiate fraudulent transactions from legitimate ones?

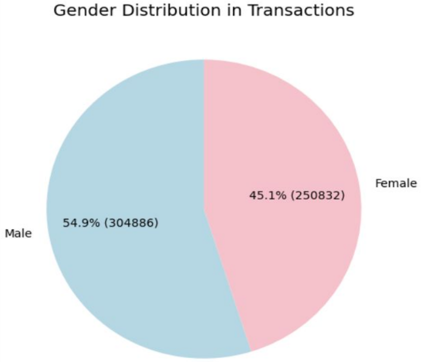


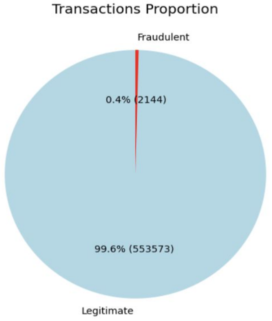
D7: Shreya, Jasmine Chan, Snowy To, Catherine Ye, Sean Hsu, Jerry Wang, Severus Zhang

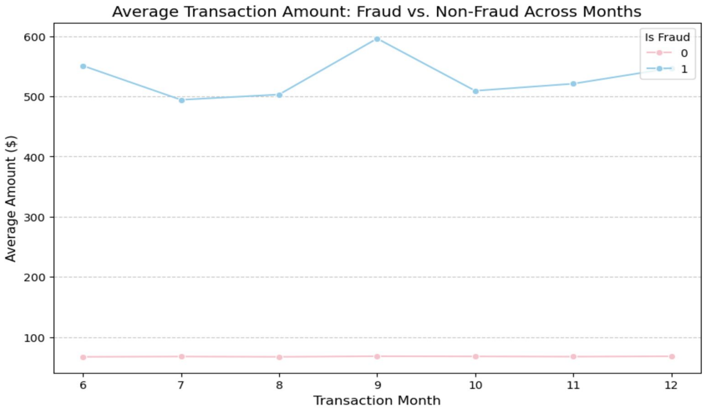
Data Overview and Key Variables

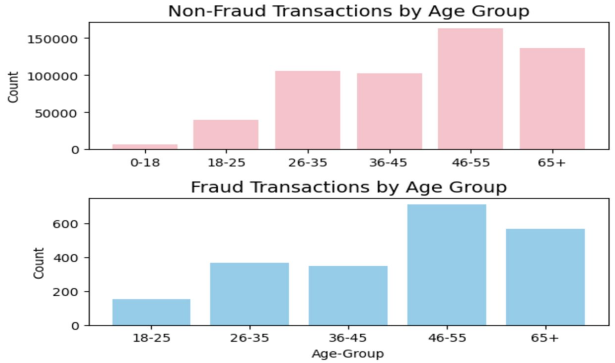












Fraud in Focus: A deep dive into anomalous activities

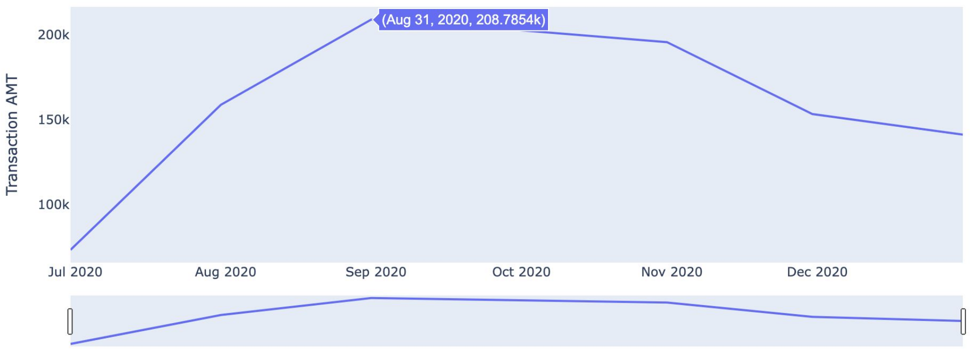
Understanding the overall transaction patterns is essential, but it's the fraudulent activities that drive costs, risk, and mitigation efforts. Let's explore this critical segment.

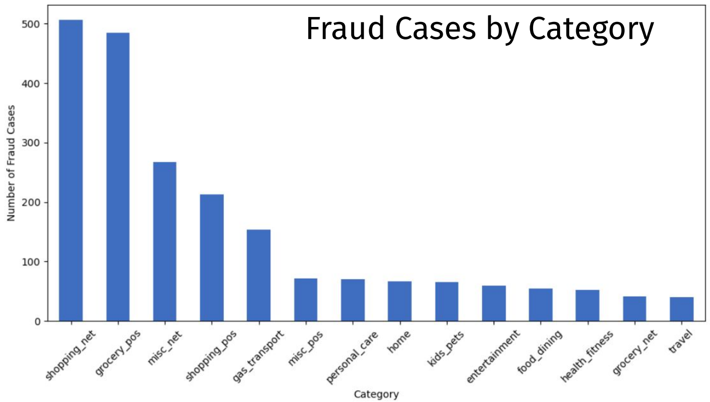
Core Analytical Dimensions:

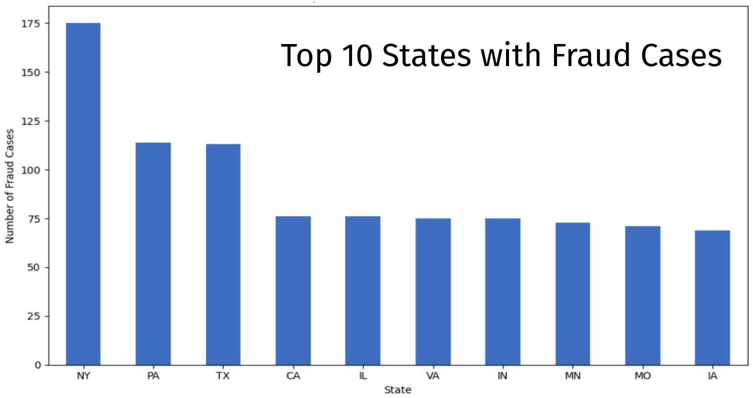
- **Transaction Month** – Analyze trends and patterns by month.
- **Spending Categories** – Break down transaction types by categories
- **Geographic Distribution** – Explore transaction patterns across states and cities.
- **Transaction Timing** – Understand the time-based dynamics of transactions.
- **Demographic Insights** – Segment transactions by age groups.
- **Spending Behavior** – Evaluate average spending per individual.
- **Weekend Effect** – Assess transaction activity on weekends versus weekdays.

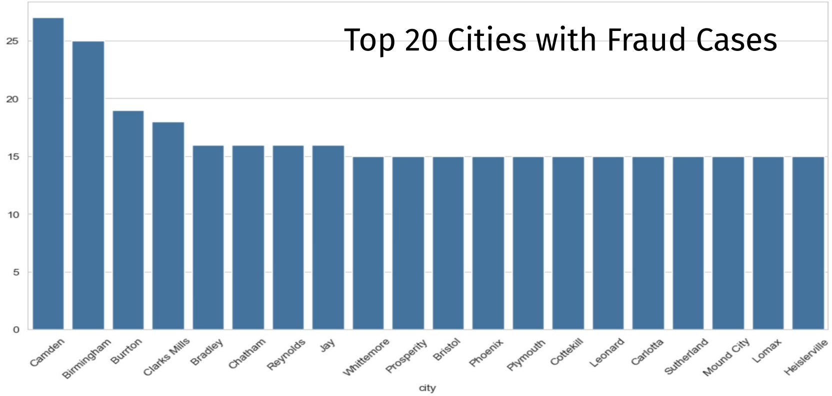
EDA: Date, Categories and Regional Insights

Fraud Amount by each Month

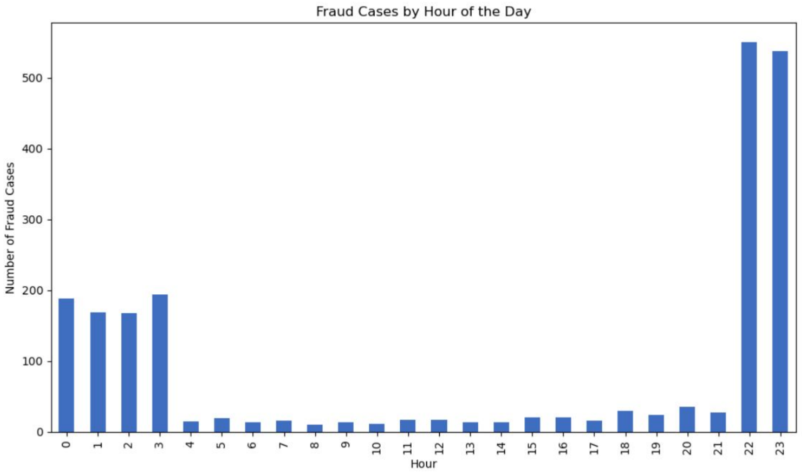


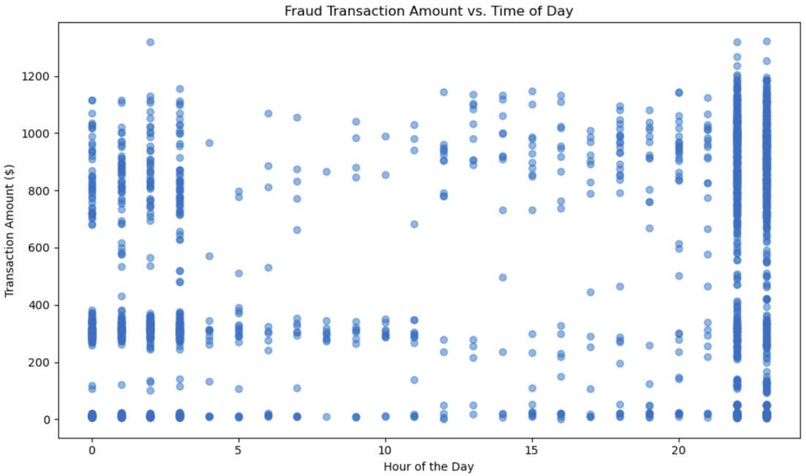


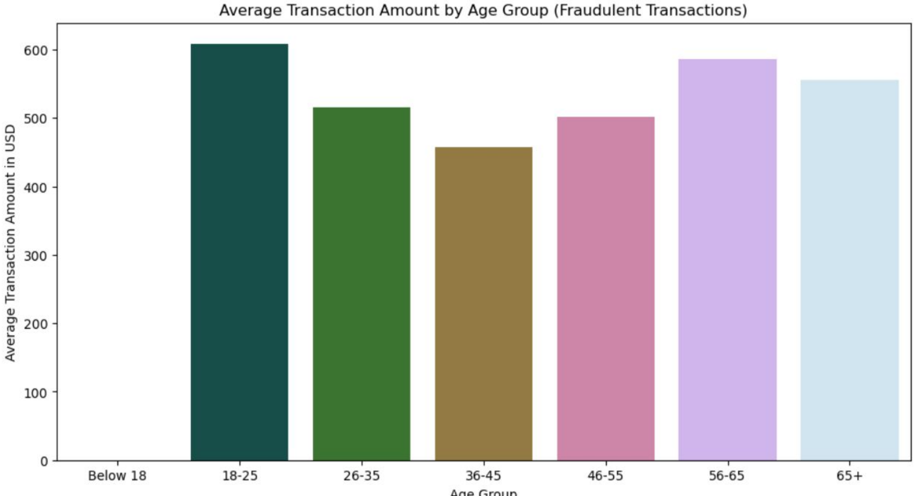


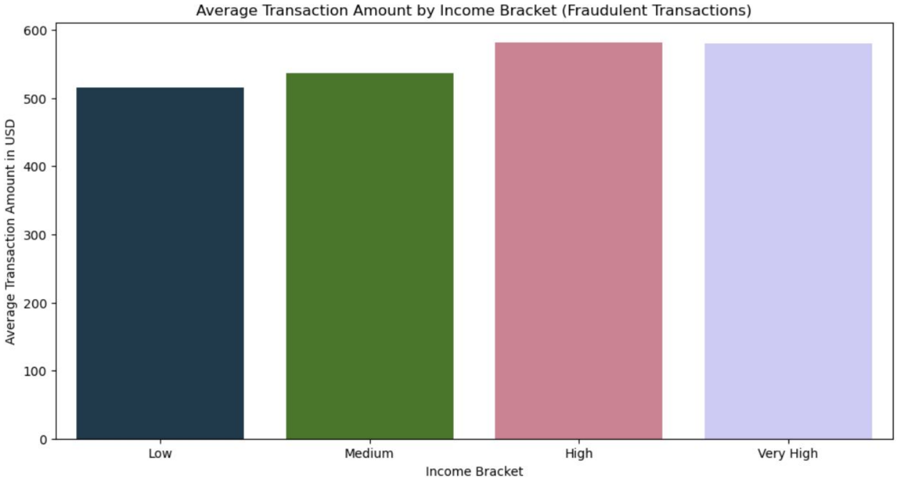


Fraud by Hour, Age Group, and Income Bracket









Hypothesis Testing and Logistic Regression

Is weekday/weekend and fraud or not independent?

```
Chi-square Statistic: 0.2867907489905829
p-value: 0.5922844718406959
Degrees of Freedom: 1
Expected Frequencies:
[[303709.18137404 249864.81862596]
 [  1176.81862596    968.18137404]]
```

Is there a statistical difference in the fraud transaction amount of male different than those of female?

```
TtestResult(statistic=6.409727639782657, pvalue=1.7873917726692276e-10, df=2143.0)
```

Logit Regression Results

Dep. Variable:	is_fraud	No. Observations:	57502
Model:	Logit	Df Residuals:	57466
Method:	MLE	Df Model:	35
Date:	Thu, 05 Dec 2024	Pseudo R-squ.:	0.4208
Time:	12:51:46	Log-Likelihood:	-5304.6
converged:	True	LL-Null:	-9158.7
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-7.0751	0.250	-28.281	0.000	-7.565	-6.585
amt	0.0061	0.000	45.904	0.000	0.006	0.006
gender_M	-0.1384	0.057	-2.407	0.016	-0.251	-0.026
age	0.0058	0.002	3.552	0.000	0.003	0.009
city_pop	-5.601e-07	1.33e-07	-4.209	0.000	-8.21e-07	-2.99e-07
distance	9.776e-05	0.001	0.101	0.920	-0.002	0.002
is_weekend	0.0642	0.061	1.045	0.296	-0.056	0.185
category_food_dining	0.2127	0.195	1.093	0.274	-0.169	0.594
category_gas_transport	2.0100	0.186	10.822	0.000	1.646	2.374
category_grocery_net	1.8286	0.231	7.927	0.000	1.376	2.281
category_grocery_pos	2.6057	0.170	15.360	0.000	2.273	2.938
category_health_fitness	0.1855	0.197	0.943	0.346	-0.200	0.571
category_home	-0.0183	0.185	-0.099	0.921	-0.381	0.344
category_kids_pets	0.0599	0.186	0.322	0.748	-0.305	0.425
category_misc_net	1.0997	0.184	5.978	0.000	0.739	1.460
category_misc_pos	0.5704	0.196	2.915	0.004	0.187	0.954
category_personal_care	0.3689	0.184	2.008	0.045	0.009	0.729
category_shopping_net	0.2060	0.178	1.157	0.247	-0.143	0.555
category_shopping_pos	-0.3513	0.184	-1.908	0.056	-0.712	0.010
category_travel	-0.8717	0.269	-3.243	0.001	-1.399	-0.345
time_of_day_morning	-0.7921	0.197	-4.026	0.000	-1.178	-0.406
time_of_day_evening	2.6764	0.142	18.826	0.000	2.398	2.955
time_of_day_midnight	1.3826	0.164	8.424	0.000	1.061	1.704
job_category_Construction	0.2236	0.158	1.418	0.156	-0.085	0.533
job_category_Consulting	0.0737	0.215	0.342	0.732	-0.348	0.496
job_category_Education	0.0159	0.152	0.105	0.917	-0.282	0.313
job_category_Engineering & Tech	0.0940	0.134	0.699	0.484	-0.169	0.357
job_category_Finance	0.0558	0.163	0.343	0.732	-0.263	0.375
job_category_Goverment	-0.1652	0.191	-0.864	0.388	-0.540	0.210
job_category_Healthcare	0.0905	0.134	0.678	0.498	-0.171	0.352
job_category_Insurance	0.2209	0.296	0.747	0.455	-0.359	0.801
job_category_Legal	-0.8499	0.306	-2.776	0.005	-1.450	-0.250
job_category_Management	-0.0144	0.205	-0.070	0.944	-0.416	0.387
job_category_Media	-0.1237	0.161	-0.767	0.443	-0.440	0.192
job_category_Retail	0.5071	0.240	2.115	0.034	0.037	0.977
job_category_Others	0.1591	0.130	1.221	0.222	-0.096	0.414

Key Business Implications

01

Target High-Risk Transactions

Focus on high-risk categories like grocery and gas purchases.

02

Time-Sensitive Monitoring

Allocate fraud detection resources during evening and midnight hours,

03

Personalized Fraud Prevention

Reduce monitoring for legal professionals and male users, who are less likely encounter fraud.

04

Enhanced Security for High-Value Transactions

Implement stricter checks for higher-value transactions to minimize financial losses.

Limitations & Future Study

Data Source

Data Timeframe

Variables

Methodology

Integration of AI and Machine Learning

Biometric Authentication & Behavioral Biometrics

Customer-Centric Fraud Analytics

Impact of Global Trends