profitLens

ProfitLens: Visualizing E-Commerce Trends

Understanding which products, regions, and discount strategies drive the highest profitability through data-driven insights.

The Business Challenge

Our Mission

Analyze e-commerce data from Kaggle to uncover how products, regions, and discounts impact sales and profit. Discover which business areas generate the most revenue and enable smarter, data-driven decisions.

Why It Matters

Companies need clear insights into profitability drivers to optimize strategies, reduce waste, and maximize returns across categories and regions.

Project Objectives

Data Preparation

Clean and organize raw data

Pattern Discovery

Identify sales, profit, and discount relationships

Profitability Analysis

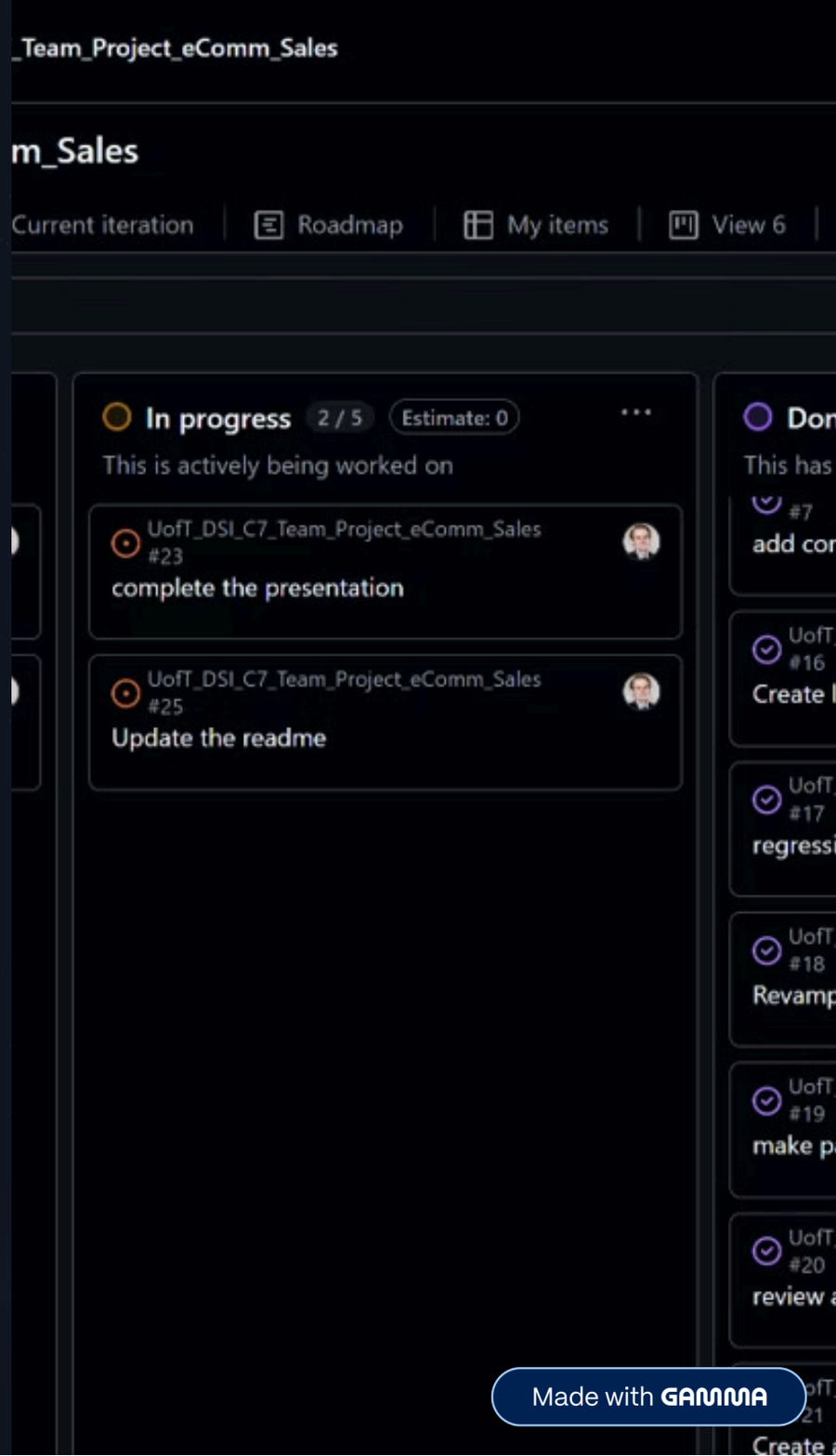
Spot most profitable categories/regions

Visual Insights

Communicate results with clear charts

Predictive Modeling

Forecast future profit and sales



Team & Responsibilities

Vikrama

Business proposal development, project naming, data cleaning, experiments, and code review

Iryna

Dataset selection, data exploration, KPI development, and setting business goals

Paul

Repository management, project tracking, data exploration and cleaning, slides, visualization, and presentation management

Key Data Parameters

1

Price

Pricing & revenue

2

Product Quantity

Volume & inventory

3

State

Regional performance

4

Repeat orders

Driving returning customers

5

Date

Temporal trends

```
# Train/test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test

# Fit regression model
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Evaluate
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print(f"\n SKU Regression Results:")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.4f}")

# Feature importance
coefficients = pd.Series(model.coef_, index=X.columns).sort_va
print("\n Top 15 Positive Features (drive higher sales):\n", c
print("\n Top 15 Negative Features (lower sales):\n", coefficie

# Find top-selling SKUs
print("\n Top 10 Most Popular SKUs:\n")
print(agg_df[['SKU', 'Category', 'Size', 'Qty']].sort_values('
```

```
2.0s

Users\Paul\AppData\Local\Temp\ipykernel_9260\99505703.py:22: Dt
 = pd.read_csv(data_path)

 Regression Results:
: 19.35
 0.0455

 15 Positive Features (drive higher sales):
egory_Western Dress     14.529545
gory_Set                7.448648
_6XL                    7.118155
gory_kurta              6.180175
```

Data Cleaning Strategy

Essential Steps

- Remove unnecessary columns (index, promotion-ids)
- Handle missing values in Courier Status, currency, Amount, ship-city, ship-state
- Convert Date to proper format for time-based analysis
- Standardize text formats to lowercase

Data Enhancement

- Check and remove duplicate Order IDs
- Group similar order statuses for consistency
- Ensure currency consistency (convert to INR)
- Add derived columns: Profit Margin, Month/Year extractions

Analysis Focus Areas

Product Insights

Top selling and highest revenue categories

Regional Insights

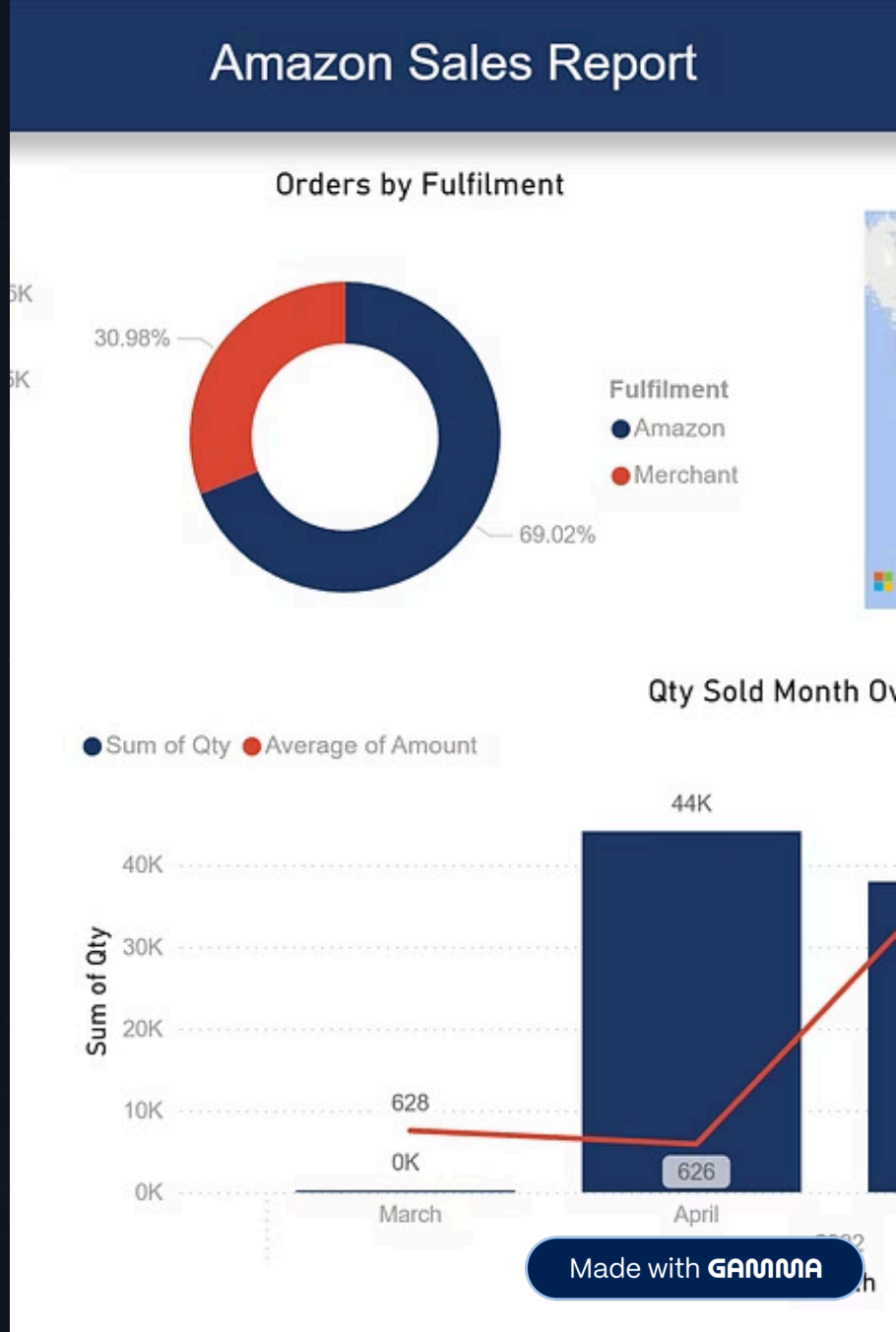
Most profitable states and cities

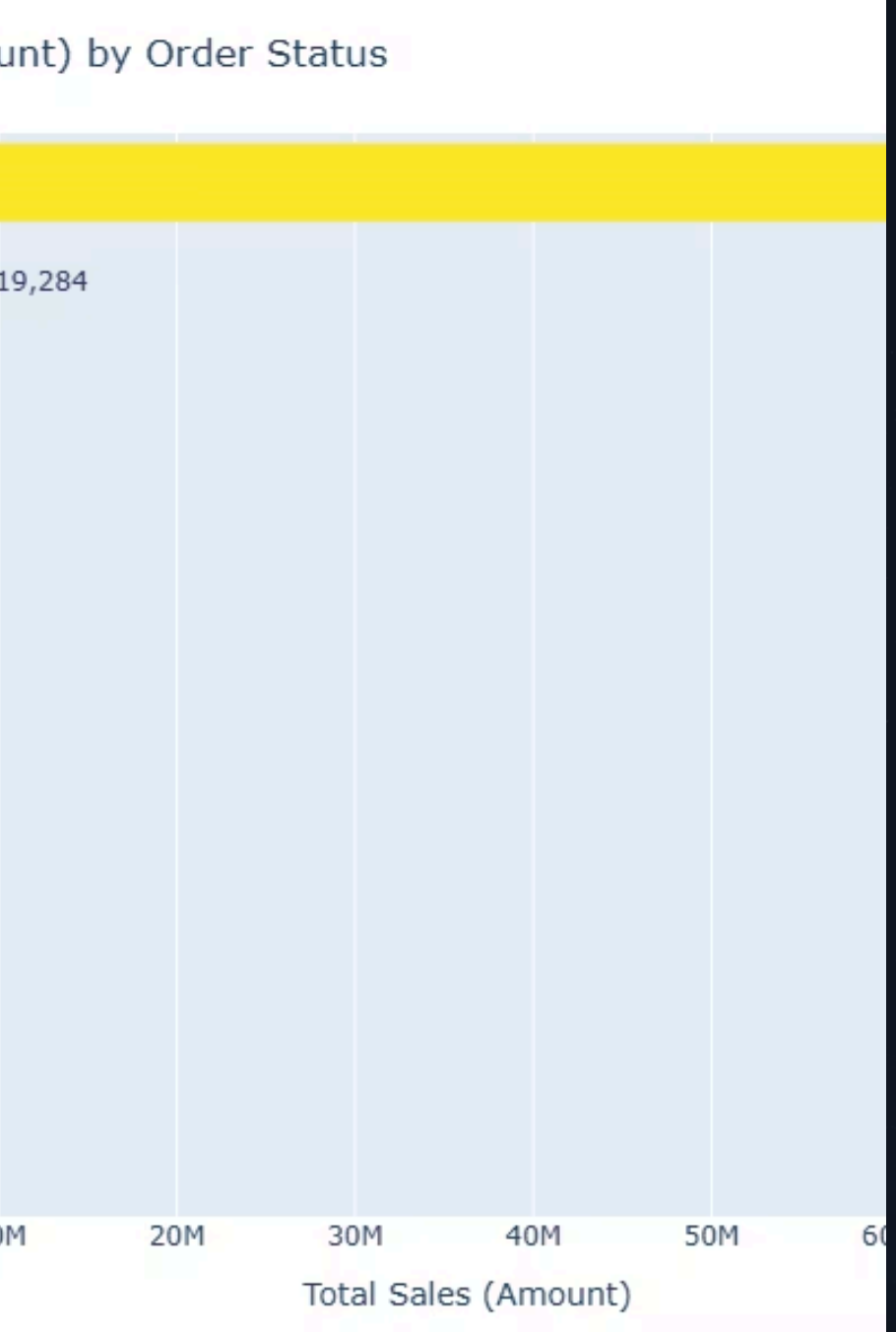
Order Performance

Amazon vs Merchant fulfillment comparison

Sales Patterns

Impact of discounts, promotions and **repeat orders**





Visualizations

- **Plotly, Seaborn, Matplotlib:** Compare product performance across types.
- **Power BI:** Track shipped, cancelled, and pending orders.
- **Monthly Sales Trends:** Identify seasonal spikes and patterns.
- **Correlation Heatmap:** Understand relationships between sales, quantity, and discounts.

Risks & Challenges

Profit Margin Calculation

No clear methodology to calculate profit margins given current dataset limitations

Cost Data Gaps

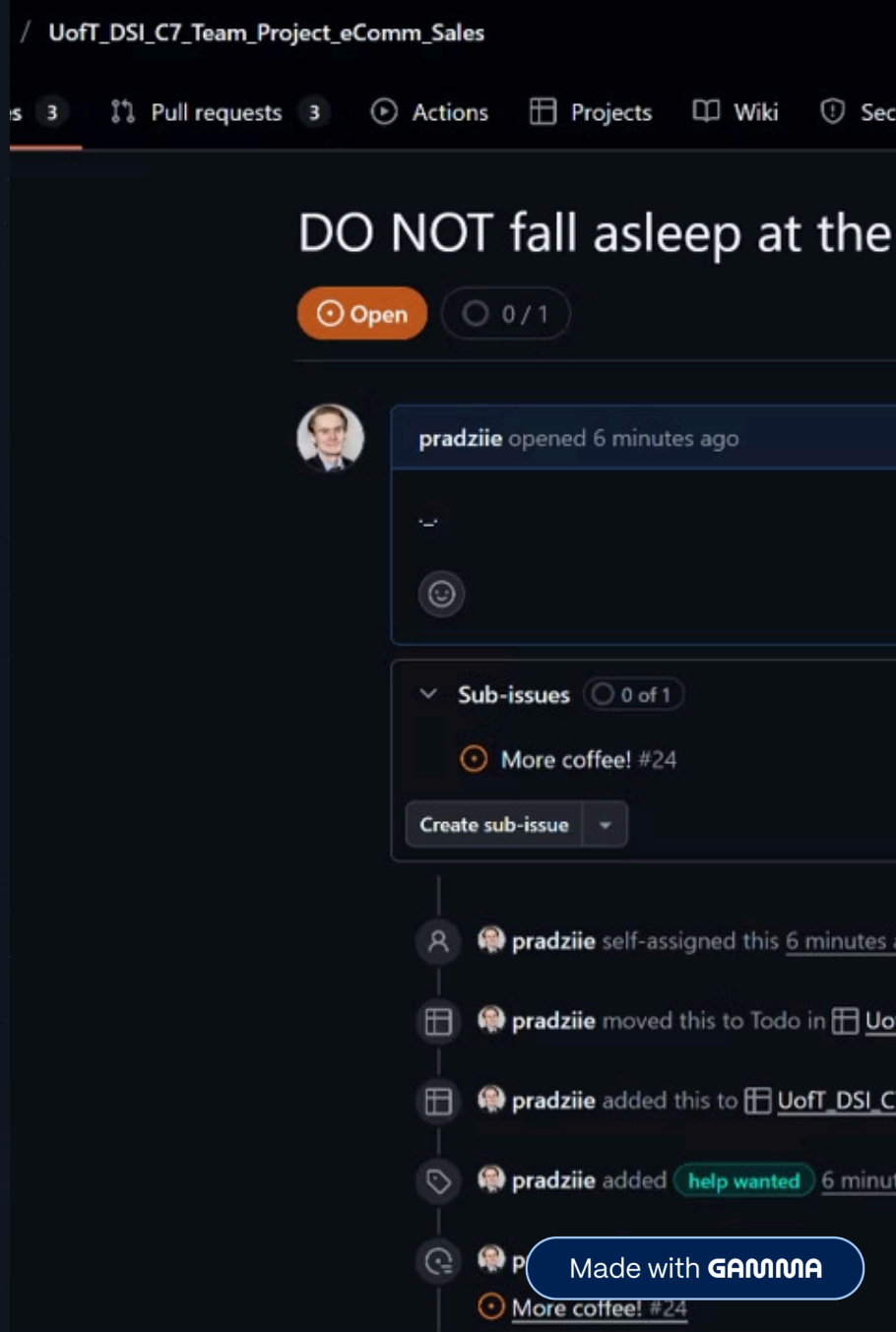
Lack of cost and Bill of Materials (BOM) data restricts deeper financial analysis

Time Constraints

Limited project timeline requires prioritization of key deliverables and analysis

PM methodology

GitHub Project used with agile methods.



Expected Outcomes

01

Clean Dataset

Organized, standardized data

02

Profitability Insights

Identify key profit drivers

03

Visual Dashboard

Interactive data insights

04

Predictive Model

Future sales forecasting

05

Actionable Recommendations

Data-driven strategies for growth

06

Bonus

UofT **color schemes** for Excel and Power BI (available in readme)