LGS Customer & Revenue Data Analytics PoC

Python Data Wrangling & Exploratory Analytics

Kanty Gakima – Data Engineer, Jarvis Consulting

o Business problem & objectives

o Dataset Overview & Data quality

o Data quality, risks & limitations

o Analytics methods & approach

o Key customer & Revenue insights

o Recommendations

o Conclusion and Q&A

Agenda

Business Problem & Objectives

o London Gift Shop revenue stagnating despite 10+ years of online operations

o Limited understanding of customer purchasing patterns

Objectives

o Analyze the transactional dataset provided by the LGS IT team to uncover meaningful purchasing patterns.

o Generate insights on customer shopping behaviors, seasonality, cancellations, and geographic trends.

Dataset Overview

o   SQL dataset containing transactions from December 2009 to September 2011

o   The file has ~1M rows and 8 columns detailing invoices, product description, quantities, pricing, customers , and countries.

o   Data was exported by LGS IT team, and anonymized

o   Loaded into a PostgreSQL database and analyzed using Python (Pandas) for data wrangling and analytics.

Data quality, risks & limitations

o **Inconsistent and incomplete records** (cancelled invoices, missing IDs) reduce reliability of customer and revenue insights.

o **Fragmented product data & short 2-year window** limits product-level analysis and long-term forecasting.

Mitigations

o Build a **Product Master Table** to standardize SKUs and enable meaningful product/category analytics.

o Improve **data capture & identification** (e.g., stable customer IDs, richer event logging) to support deeper future analysis.

Data cleaning & transformations

o   **Type casting**

o   **Derived fields created**

o   **Filtering and normalization**

o   **Standardization of column naming and formats**
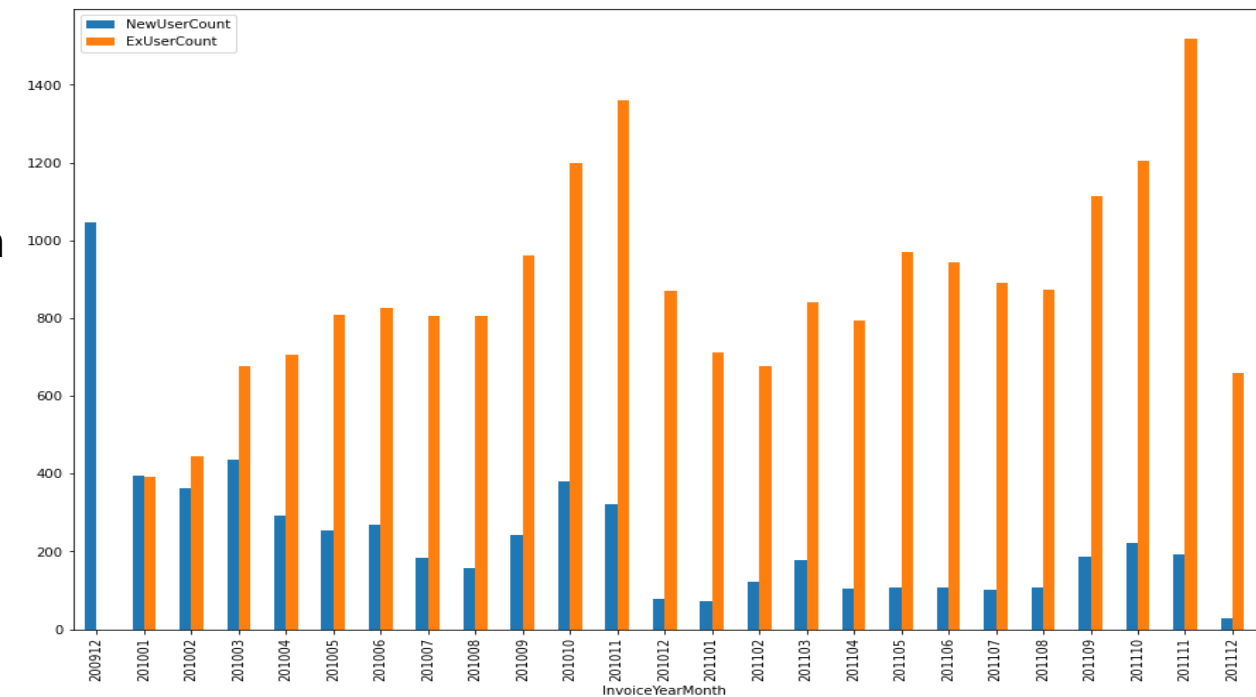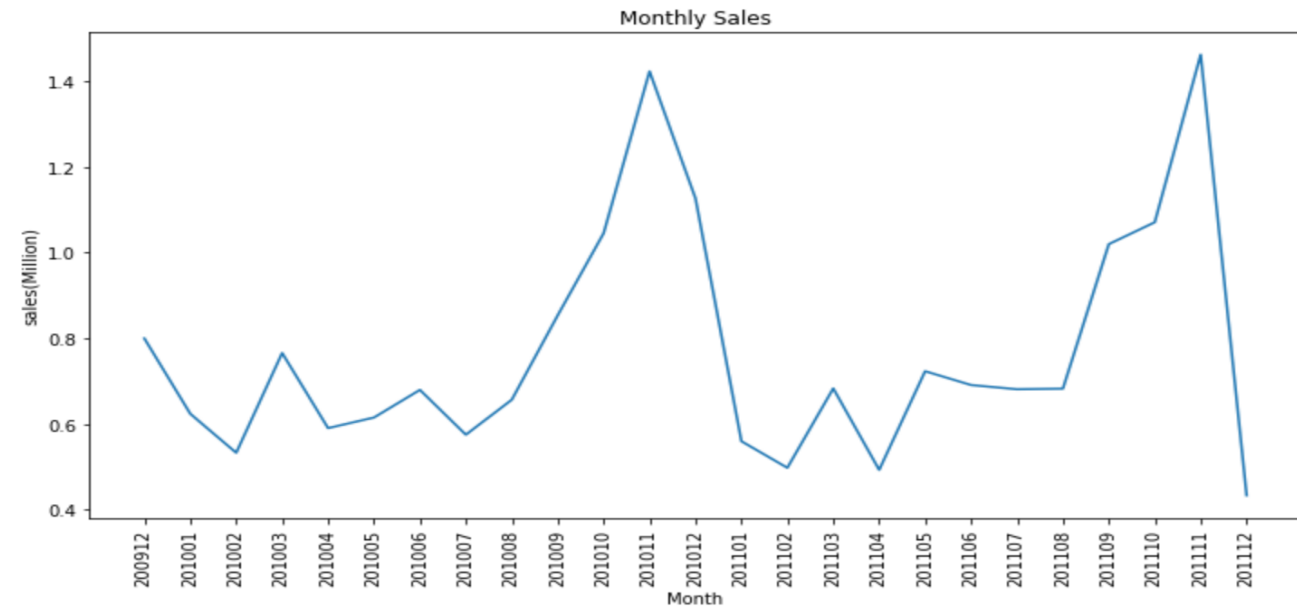
Analytics methods & approach

o **Time-series analysis** to evaluate monthly sales, customer activity, and cancellation trends, revealing seasonality and operational gaps.

o **Ranking and aggregation techniques** to identify top customers, top countries, and product performance patterns.

o Performed **RFM segmentation** (Recency, Frequency, Monetary) to classify customers into actionable behavioral groups.

o **Customer lifecycle analysis** by distinguishing first-time vs repeat buyers.

Insight 1 : Customer growth and seasonality

Business Question: How does customer purchasing behavior evolve throughout the year, and how can LGS better leverage both new and repeat customers?

Recommendation / Business Impact:

o Plan early for November peaks (sales up 2–3X) by boosting inventory 2–3 months ahead.

o Run targeted seasonal campaigns in Sept–Nov, when new customer acquisition rises.

o Strengthen loyalty programs, since 75–90% of monthly buyers are repeat customers.

o Stabilize off-season demand with retention tactics and off-season promos.
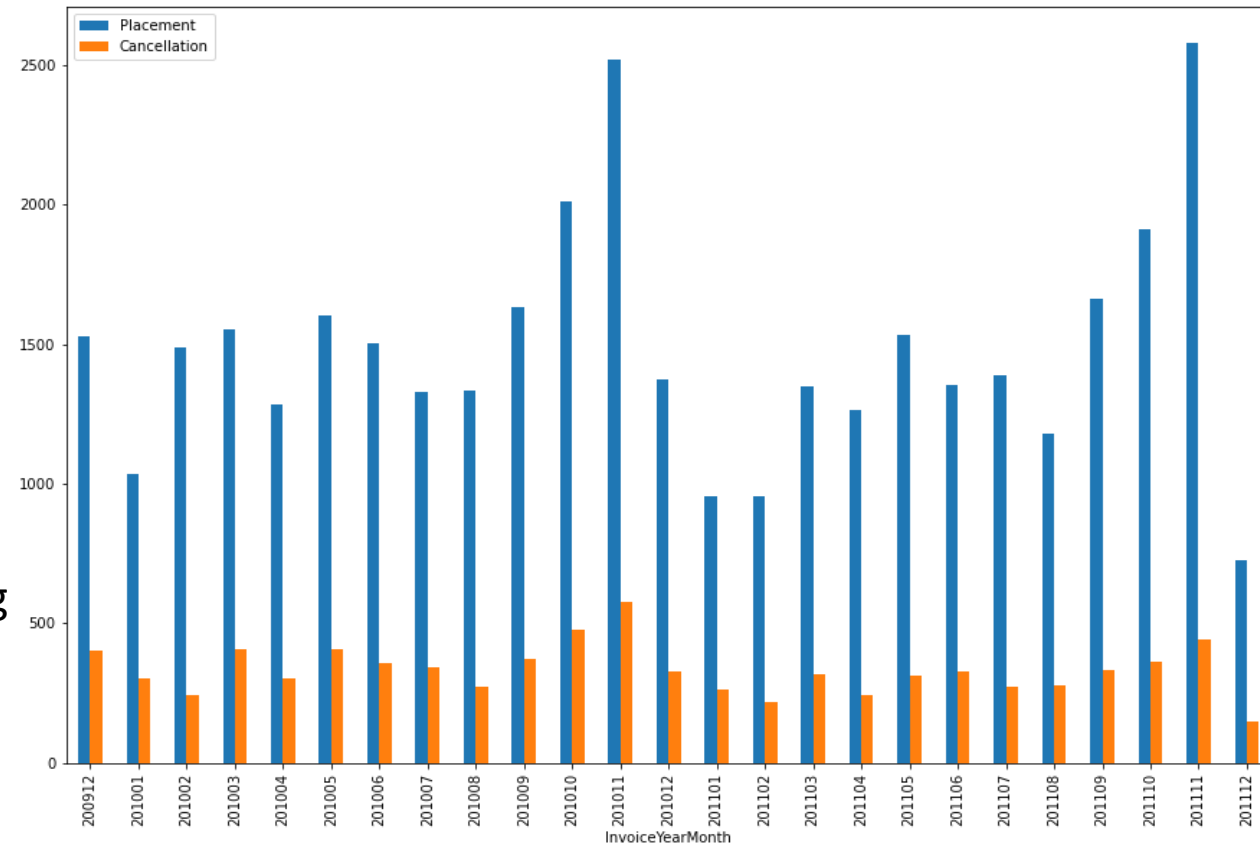
Insight 2 : Order Cancellations trends (Placed vs Cancelled orders)

Business Question: How significant are cancellations, and do they impact LGS's revenue?

Recommendation / Business Impact:
o Investigate root causes of cancellations (stockouts, listing errors, fulfillment issues).

o Strengthen inventory management and demand planning Before seasonal peaks.

o Reducing cancellations by 5 – 10% would recover meaningful revenue and improve customer satisfaction.

Insight 3 : RFM Segmentation for customers categorization

Analysis method that groups customers based on their recency, frequency, and monetary value; scoring customers on these metrics to categorize them and conduct tailored campaigns

Recommendation / Business Impact:
- **Champions:** Provide exclusive offers, early access, and loyalty perks, improves retention & lifetime value.
- **Can't Lose:** Re-engage immediately through personalized campaigns, replenishment reminders, and discounts.
- **Hibernating:** Win-back campaigns focusing on past purchases and incentives to revive dormant customers.

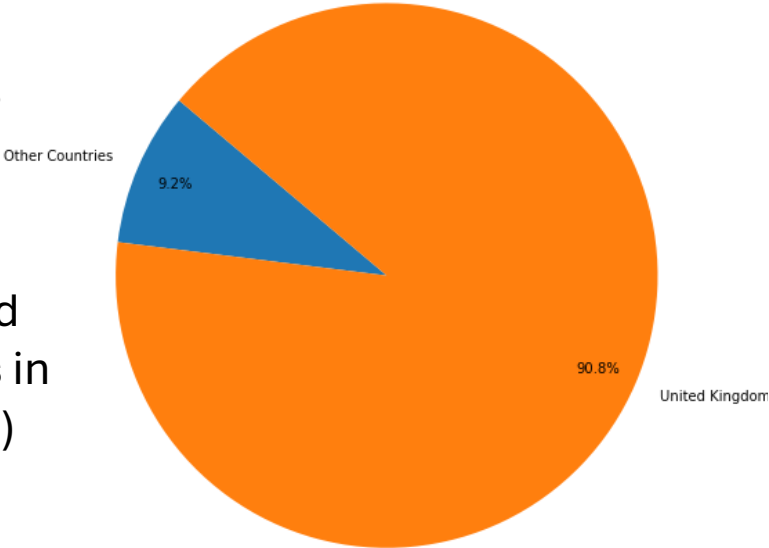| Segment	Recency mean	Recency count	Frequency mean	Frequency count	Monetary mean	Monetary count
About to Sleep	5222.689474	380	1.460526	380	492.810900	380
At Risk	5491.094415	752	4.672872	752	1157.449162	752
Can't Lose	5436.305882	85	17.717647	85	5725.638494	85
Champions	5122.512850	856	23.532710	856	10510.937634	856
Hibernating	5580.237785	1535	1.336808	1535	340.030536	1535
Loyal Customers	5182.149565	1150	11.941739	1150	3966.024762	1150
Need Attention	5228.395604	273	3.673993	273	1099.877403	273
New Customers	5124.814815	54	1.000000	54	362.463148	54
Potential Loyalists	5139.674764	741	2.967611	741	899.359865	741
Promising	5152.758621	116	1.000000	116	322.758966	116

Insight 4: Geographic Market Concentration & Expansion opportunity

Business Question: Which countries contribute the most customers, and how geographically concentrated is LGS's customer base?

Recommendation / Business Impact:
o **Invest in targeted digital advertising** and localized seasonal promotions to increase brand awareness in high-potential close regions(Germany, France, etc.)
o **Optimize shipping/logistics options**

o **Develop geo-specific email campaigns** and product recommendations tailored to local preferences.

Customer Distribution by Geography

Other Countries
9.2%

90.8%

United Kingdom

10 first other countries with the most customers

	Country	UniqueCustomers	Percentage
0	Germany	107	1.80
1	France	95	1.60
2	Spain	41	0.69
3	Belgium	29	0.49
4	Portugal	24	0.40
5	Netherlands	23	0.39
6	Switzerland	22	0.37
7	Sweden	19	0.32
8	Italy	17	0.29
9	Australia	15	0.25

Recommendations

o **Strengthen customer retention** by activating RFM segments ,through targeted promotions, automated win-back campaigns, and personalized recommendations.

o **Prepare for seasonal demand** by optimizing inventory, improving forecast accuracy, and launching early marketing campaigns ahead of holiday spikes.

o **Reduce cancellations** through better product data quality, improved stock accuracy, and proactive communication workflows.

o **Expand internationally** by prioritizing geographically close, high-potential markets with localized advertising and optimized logistics.

o **Build scalable analytics foundations** by automating ETL flows, standardizing product data, and gradually transitioning toward real-time dashboards and distributed processing (Spark/Databricks).

THANK YOU

Jarvis

EQUALIZING OPPORTUNITY ——— MAXIMIZING POTENTIAL

JRVS.CA