📊Customer Lifetime Value, Cohort & Funnel Analysis (Python)

📌 Project Overview

This project focuses on understanding customer behavior and business performance by combining **Customer Lifetime Value (CLV)** analysis, **cohort-based retention analysis**, and **funnel conversion analysis**.

The objective is to identify high-value customers, analyze retention trends over time, and detect conversion drop-offs across the customer journey, enabling data-driven marketing, retention, and growth strategies.

🎯 Objectives

- ★ Estimate Customer Lifetime Value (CLV)
- ★ Segment customers based on value contribution
- ★ Analyze customer retention using cohort analysis
- ★ Identify funnel drop-offs and conversion rates
- ★ Deliver actionable datasets for business stakeholders

🛠️ Tools & Technologies

- ★ **Language:** Python
- ★ **Environment:** Google Colab (Cloud-based Jupyter Notebook)
- ★ **Libraries Used:**
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn

📊 Dataset

- ★ **Type:** Synthetic but realistic customer transaction dataset
- ★ **Nature:** Raw and uncleaned (real-world inspired)
- ★ **Dataset Characteristics**
  - Multiple customer interactions
  - Missing signup dates
  - Negative revenue values (refunds)
  - One-time and repeat customers
  - Multiple acquisition channels
  - Funnel-stage level events

The dataset was intentionally designed to mimic **real production data** that requires business-rule-based cleaning.

🧹 Data Cleaning & Assumptions

Key assumptions applied:

- ★ Missing signup dates replaced with first observed order date
- ★ Negative revenue treated as refunds and excluded from CLV
- ★ CLV calculated using only purchase-related events
- ★ Customer lifetime defined as time between first and last purchase

These assumptions were explicitly documented before cleaning to ensure transparency and reproducibility.

💰 Customer Lifetime Value (CLV) Analysis

- ★ CLV distribution is highly right-skewed
- ★ A small percentage of customers generate most of the revenue
- ★ Customers were segmented into Low, Mid, High, and VIP value groups
- ★ Purchase frequency showed a stronger relationship with CLV than customer lifespan

---

📊 Cohort Analysis (Customer Retention)

- ★ Customers were grouped into monthly signup cohorts
- ★ Retention rates were tracked over time
- ★ Significant churn occurs after the first month
- ★ Strong early retention is closely linked to higher lifetime value

A cohort heatmap was used to visualize retention decay patterns across time.

🔻 Funnel Analysis (Conversion & Drop-offs)

- ★ Funnel stages analyzed: Visit → Signup → Purchase → Repeat Purchase
- ★ Major drop-offs observed in early funnel stages
- ★ Repeat purchase conversion was relatively low
- ★ Channel-wise funnel analysis revealed differences in acquisition quality

---

🧠 Advanced Insights

- ★ Signup cohorts with higher early retention consistently generated higher CLV
- ★ Customers reaching repeat purchase stages contributed significantly more revenue
- ★ High-volume acquisition channels were not always high-value channels
- ★ Early lifecycle engagement has a compounding impact on long-term revenue

---

📤 Business Deliverables

The project generated multiple **business-ready datasets**:

- ★ **High-value customers list** for marketing and loyalty teams
- ★ **Retention-risk customers list** for churn prevention
- ★ **Channel-level CLV performance summary** for strategy teams
- ★ **Funnel conversion summary** for product and UX optimization

All deliverables were exported as CSV files for immediate stakeholder use.

📌 Conclusion

This project demonstrates how combining CLV, cohort retention, and funnel analysis provides a holistic view of customer behavior and business performance. By translating raw transactional data into actionable insights and deliverables, the analysis supports informed decision-making across marketing, retention, and growth initiatives.

🚀 Future Enhancements

- ★ Predictive CLV modeling
- ★ Survival analysis for churn prediction
- ★ Channel ROI optimization
- ★ Integration with marketing automation tools
- ★ Interactive dashboards (Power BI / Tableau)

☁️ Execution Environment

The entire project was developed and executed using **Google Colab**, ensuring cloud-based accessibility and reproducibility.

🧒 Author

Anirudha Das
Aspiring Data Analyst / Business Intelligence Professional
📍Bolpur, WB, India
