🔎 Business Problem Statement & Project Deliverables

1. Business Problem

A leading retail organization is observing shifts in customer purchasing patterns across demographics, product categories and sales channels (online vs offline). Management needs to understand what drives purchase decisions and repeat buying — for example, the effect of discounts, product reviews, seasonality, and preferred payment methods — so they can increase sales, improve customer satisfaction, and grow long-term loyalty.

Overarching question:

How can the company leverage consumer shopping data to identify trends, segment customers, and optimize marketing & product strategies to increase engagement, retention, and revenue?

---

2. Project Objectives

1. **Profile customers**: Identify key customer segments (e.g., frequent buyers, churn risk, high-value customers) and understand their behavioral patterns.
2. **Uncover purchase drivers**: Quantify the impact of promotions, reviews, seasonality, payment choices and channels on conversion and repeat purchase.
3. **Measure channel performance**: Compare online vs offline sales performance across products, times, and customer segments.
4. **Prioritize interventions**: Recommend actionable marketing, merchandising and product strategies to increase conversion and retention.
5. **Operationalize insights**: Deliver an interactive dashboard and reproducible analysis pipeline so stakeholders can monitor trends continuously.

---

3. Scope (In-Scope / Out-of-Scope)

In-Scope

- Exploratory data analysis and data cleaning of transactional and customer datasets.
- Feature engineering (recency, frequency, monetary, discount sensitivity, review sentiment proxies, seasonal indicators).
- SQL queries to derive business metrics and cohort/segmentation analysis.
- Interactive Power BI dashboard with KPIs, segment views, trend charts and drilldowns.
- A short report & presentation with findings and prioritized recommendations.

- Deliverable code/scripts (Python notebooks, SQL files, PBIX) and synthetic data if production data is unavailable.

Out-of-Scope

- Building a live production recommender or real-time pipeline (unless requested).
- Direct changes to production pricing or campaign engines.
- Customer PII re-identification or any activity that violates privacy rules.

---

4. Success Criteria / KPIs

Measured at project close (baseline → post-project):

- **Analysis completeness**: All core questions answered (segments identified, drivers quantified).
- **Actionability**: ≥3 prioritized recommendations with expected impact and justified by data.
- **Stakeholder acceptance**: Dashboard and report approved by product/marketing lead.
- **Reproducibility**: Analysis scripts and SQL run successfully on sample dataset.
- **Clarity**: Clear visualizations and a 1-page summary that non-technical stakeholders understand.

Business KPIs to track as a result of recommendations (examples):

- Increase in repeat purchase rate (%)
- Increase in average order value (AOV)
- Reduction in lost bookings / abandoned carts (%)
- Improvement in retention rate of targeted segments (%)

---

5. Deliverables (detailed)

1. **Data Preparation & Modeling (Python)**
    - Jupyter notebooks / Python scripts to clean raw data, handle missing values, normalize fields, create time/date features, and produce analytical tables.
    - Output: `data/processed/*.csv`, notebook `notebooks/data_prep.ipynb`.
2. **Data Analysis (SQL)**
    - Schema design for analytical tables; SQL scripts for ETL simulations and queries that compute transactions, cohorts, RFM, conversion funnels, promo lifts.
    - Output: `sql/` folder with `schema.sql` + `queries/` (segmentation, retention, channel performance).
3. **Visualization & Insights (Power BI)**
    - Interactive Power BI report (`.pbix`) showing: Overview KPIs, Customer Segmentation, Promotion & Channel analysis, Product-level trends, Cohort retention charts, and Drill-down pages.

    - Screenshots and an exported PDF for quick review.
    - Output: `dashboard/Uber_Analytics.pbix`, `dashboard/Uber_Analytics.pdf`, `Images/*`.
4. **Forecast / Sensitivity Analysis (optional / advanced)**
    - Short time-series or uplift analysis to estimate effect of promotions or seasonality on short-term sales. Implemented in Python (Prophet/ARIMA) if requested.
    - Output: `models/forecasting.ipynb` and sample forecasts.
5. **Report & Presentation**
    - Concise written report summarizing approach, key findings, recommended actions (prioritized), and estimated impact.
    - Slide deck (10–12 slides) for stakeholder presentation.
    - Output: `docs/report.pdf`, `docs/presentation.pptx`.
6. **Reproducibility & Repo**
    - Well-structured GitHub repo containing data, notebooks, SQL scripts, PBIX file, README and a short `resume_bullets.md` with copy-paste lines for your CV.
    - Docker / environment requirements in `requirements.txt` / `environment.yml` for reproducibility.

---

6. Data Sources & Assumptions

Primary data (expected):

- Transaction records (order_id, customer_id, timestamp, product_id, price, discount, payment_method, channel, status).
- Customer master (demographics, registration date).
- Product master (category, price tiers).
- Reviews / ratings (if available) or review proxies.
- Optional mobility/footfall / offline store logs (if provided).

Assumptions

- Data is pseudonymized (no PII used in analysis outputs).
- If keys or certain sources are unavailable, synthetic sample data will be generated to demonstrate pipeline.
- Timezone normalized to local business timezone (Asia/Kolkata) for all timestamps.

---

7. Constraints & Risks

- **Data quality**: Missing/incorrect timestamps or inconsistent product IDs may limit some analyses.
- **Privacy**: Must not expose PII; ensure compliance with company policies.
- **Data access**: External APIs or production datasets may have rate limits or require credentials.

- **Time**: Depth of modeling (e.g., uplift modeling) depends on data availability and time constraints.

---

8. High-level Approach & Timeline (suggested, 1–2 week sprint)

Day 1–2: Data ingestion & cleaning (Python) — produce processed tables.
Day 3: Exploratory Data Analysis + key SQL queries (cohorts, RFM, channel splits).
Day 4: Feature engineering & scoring (customer segments, promo sensitivity).
Day 5: Power BI dashboard development (KPIs, overview, segment pages).
Day 6: Report writing & insights; prepare slides.
Day 7: QA, reproducibility checks, finalize README, and deliver.

Adjust scope for additional forecasting or deeper modeling.

---

9. Roles & Responsibilities (example)

- **Data Analyst (You)**: Data cleaning, SQL queries, feature engineering, dashboard building, presenting insights.
- **Product / Marketing Owner**: Provide domain context, validate business KPIs, approve recommendations.
- **Data Engineer (optional)**: Help with data extraction or DB access if needed.

---

10. Expected Output Files (repo structure)

```
/data/
  raw/
  processed/
/notebooks/
  data_prep.ipynb
  eda.ipynb
/sql/
  schema.sql
  queries/
    rfm.sql
    retention.sql
/models/
  forecasting.ipynb
/dashboard/
  Customer_Trends.pbix
  Customer_Trends.pdf
/docs/
  report.pdf
  presentation.pptx
README.md
requirements.txt
```

11. Final Note

This project delivers a reproducible analytics pipeline and an interactive dashboard to support data-driven marketing and product decisions. The outputs are designed to be actionable: each recommended intervention is backed by quantified insights and an estimated business impact, helping stakeholders prioritize efforts that improve customer engagement, retention and revenue.