

### Prompt 1 — Full Exploratory Data Analysis (EDA) with Actionable Insights

"You are a **Senior Data Analyst** with expertise in Python (Pandas, NumPy, Matplotlib, Seaborn) and data storytelling. I have a dataset in CSV format containing sales data for an Indian retail chain (2018–2024) with 50,000 rows and 12 columns (date, location, category, units sold, price, discount, etc.).

Your task:

1. Import and inspect the dataset to understand structure, data types, and column meanings.
2. Generate summary statistics for both numerical and categorical columns, highlighting anomalies.
3. Detect missing values, quantify them column-wise, and suggest domain-specific imputation or removal strategies.
4. Identify outliers using both statistical (Z-score, IQR) and visual (boxplots) methods, explaining potential business causes.
5. Create a correlation heatmap for numerical features and explain the top 5 strongest relationships in simple business terms.
6. Provide at least 3 actionable business insights and possible next analytical steps.

**Output format:** A Jupyter Notebook with fully commented code, inline plots, and a concluding Markdown section explaining insights in non-technical language."

### Prompt 2 — Interactive KPI Dashboard for Decision-Makers

"You are a **Business Intelligence (BI) Dashboard Expert** skilled in Power BI, Tableau, and Google Data Studio. I have quarterly sales data for multiple product categories in India for the last 5 years.

A SECTION 8 – NON-PROFIT ORGANIZATION REGISTERED UNDER THE COMPANIES ACT, 2013

Your task:

1. Identify 5–7 key performance indicators (KPIs) relevant to retail business health (e.g., total revenue, gross margin, average order value, conversion rate).
2. Design an interactive dashboard layout showing KPIs as cards at the top, trend charts in the middle, and filters (by time, category, location) on the side.
3. Include drill-down capability so a user can click on a KPI and view detailed breakdowns by category, region, or month.
4. Add a geographic heatmap showing sales distribution across states, with hover tooltips.
5. Include an export-to-PDF function for monthly reporting.

**Output format:** Provide a step-by-step dashboard build guide (tool-agnostic), a mockup diagram of the dashboard, and sample formulas for KPI calculations."

### Prompt 3 — Predictive Sales Forecasting with Model Comparison

"You are a **Data Scientist** specialising in forecasting. I have monthly sales data for an e-commerce platform from January 2018 to December 2024.

Your task:

1. Perform time-series decomposition to analyse trend, seasonality, and residual components.
2. Build at least two forecasting models (ARIMA/SARIMA and Facebook Prophet).
3. Compare model performance using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error).
4. Plot actual vs predicted sales for both models and highlight differences.
5. Provide recommendations for which model to deploy, along with a 12-month sales forecast.

**Output format:** A Python Jupyter Notebook with all code, plots, and a Markdown cell comparing models with business-friendly explanations."

#### Prompt 4 — Automated Data Cleaning and Preprocessing Script

"You are a **Data Preprocessing Automation Specialist** skilled in Python and Pandas. I have a CSV file containing a mix of numerical, categorical, and datetime fields, with missing values and inconsistent formats.

Your task:

1. Write a reusable Python script to detect and handle missing values using mean/median/mode or forward-fill/backward-fill depending on the column type.
2. Remove duplicate rows and flag near-duplicates for manual review.
3. Normalise numerical columns (min-max or z-score scaling) and encode categorical columns (label or one-hot encoding as appropriate).
4. Convert date columns to proper datetime format and extract features (day, month, year, day-of-week).
5. Save the cleaned dataset to a new CSV file with a timestamped filename.

**Output format:** A fully commented Python script that can be reused for different datasets, with clear function definitions."

#### Prompt 5 — Business Data Storytelling for Stakeholder Reports

"You are a **Business Data Storyteller** with experience in creating executive summaries from analytical results. I have analysed customer purchase behaviour for my online store and want to present findings to the leadership team."

Your task:

1. Frame the analysis as a story — starting with the problem, key findings, and implications.
2. Select only the 5–7 most impactful visuals from the analysis, ensuring they are simple and easy to understand.
3. Explain each chart in 1–2 sentences highlighting what matters for the business.
4. Conclude with 3 actionable recommendations, each tied to a business outcome (e.g., revenue growth, cost saving).
5. Create a 2-slide PowerPoint layout that can be used in a leadership meeting.

**Output format:** A concise text storyboard + slide content that a non-technical executive can understand at a glance."

#### Prompt 6 — Real-Time Data Monitoring and Alerts

"You are a **Real-Time Data Monitoring Specialist** skilled in tools like Grafana, Kibana, and Power BI Streaming Dataflows. I operate a logistics company with live GPS and delivery data flowing in every 15 seconds.

Your task:

1. Design a real-time dashboard that displays vehicle location, delivery status, and delays in near real-time.
2. Implement colour-coded alerts for deliveries delayed beyond SLA (Service Level Agreement) thresholds.
3. Add trend visualisations for daily delivery count, average delivery time, and % on-time rate.
4. Integrate automated alerts via email and SMS for key managers when KPIs cross thresholds.

5. Ensure the system can handle data spikes (e.g., festival season).

**Output format:** A visual architecture diagram + tool integration plan + example SQL queries for alert generation."

### Prompt 7 — Sentiment Analysis of Customer Feedback

"You are an **NLP (Natural Language Processing) Specialist** with expertise in Python libraries like NLTK, SpaCy, and Transformers. I have 50,000 customer reviews collected over 2 years.

Your task:

1. Clean and preprocess the text (remove stopwords, lemmatise, handle emojis).
2. Classify sentiment into positive, neutral, and negative categories using a pre-trained BERT model.
3. Create visualisations:
  - Sentiment distribution pie chart.
  - Monthly sentiment trend line chart.
  - Word cloud for each sentiment category.
4. Identify top 5 positive and top 5 negative themes with example reviews.
5. Provide actionable recommendations for product/service improvement based on sentiment patterns.

**Output format:** A Jupyter Notebook with code, charts, and a Markdown insights summary."

### Prompt 8 — Comparative Category Performance Report

"You are a **Business Performance Analyst**. I have category-wise sales data for 10 product categories over the last 3 years.

Your task:

1. Calculate YoY (Year-over-Year) and MoM (Month-over-Month) growth rates for each category.
2. Rank categories based on revenue, profit margin, and units sold.
3. Create a dashboard view showing category trends side-by-side.
4. Highlight top 3 performing categories and bottom 3 lagging categories.
5. Suggest category-level actions to boost sales and margins for underperformers.

**Output format:** A comparative analysis table + dashboard layout mockup + 1-page action plan."

### Prompt 9 — Correlation and Causation Testing

"You are a **Data Scientist** with a focus on statistical inference. I have a dataset on marketing spend (TV, social media, influencer, print) and corresponding sales figures.

Your task:

1. Calculate correlation coefficients for each marketing channel vs sales.
2. Perform hypothesis testing to check statistical significance (p-values).
3. Run a multiple regression analysis to see which channels predict sales best.
4. Visualise results using scatter plots and regression lines.
5. Provide a plain-language explanation of findings for non-technical stakeholders.

**Output format:** Jupyter Notebook with plots + regression output table + simplified insights brief."

### Prompt 10 — Customer Churn Prediction and Retention Strategy

"You are a **Customer Analytics Expert** specialising in churn modelling. I have SaaS customer data including sign-up date, usage frequency, support tickets, and payment history.

Your task:

1. Define churn for my business context (e.g., inactive for 60 days).
2. Engineer predictive features from usage and payment history.
3. Build a classification model (Logistic Regression, Random Forest, or XGBoost) to predict churn probability.
4. Evaluate using accuracy, precision, recall, and ROC-AUC.
5. Suggest retention strategies for the top 20% at-risk customers.

**Output format:** Python Notebook with code + confusion matrix + strategic retention plan."

### Prompt 11 — Data Visualization Best Practices Guide

"You are a **Data Visualization Trainer**. Prepare a best practices guide for visualising financial performance data for stakeholders.

Your task:

1. Recommend which chart types to use for time-series, category comparison, and part-to-whole analysis.
2. Suggest an accessible, colour-blind-friendly palette.
3. Explain how to avoid misleading scales and data distortion.
4. Include 3 examples of excellent visualisations and explain why they work.
5. Provide 3 poor visualisation examples and show corrected versions.

**Output format:** A 5-page PDF guide with do's and don'ts + visual examples."

### Prompt 12 — Multi-Dataset Integration Workflow

"You are a **Data Integration Specialist** skilled in ETL (Extract, Transform, Load) processes. I have three datasets:

- Customer demographics (Excel)
- Purchase history (CSV)
- Web analytics data (Google Analytics export)

Your task:

1. Identify common keys for merging datasets.
2. Clean and standardise column formats and naming.
3. Join datasets into a master table.
4. Perform initial descriptive analysis on combined data.
5. Suggest 3 insights achievable only after combining data.

**Output format:** Python Notebook with ETL code + final merged dataset snapshot + insight summary."

### Prompt 13 — Interactive Geo-Spatial Sales Mapping

"You are a **GIS (Geographic Information Systems) Analyst**. I have state-wise sales data for India for the past 12 months.

Your task:

1. Create an interactive map showing sales density using a colour gradient.
2. Add filters for month, product category, and sales rep.
3. Display state-level tooltips with key KPIs (revenue, units sold, growth rate).
4. Enable comparison mode for two selected states.
5. Provide export options (PNG, PDF).

**Output format:** Dashboard implementation guide + sample data visualisation screenshot."

#### Prompt 14 — Industry Benchmark Comparison with Gap Analysis

"You are a **Market Intelligence Analyst** specialising in competitive benchmarking. I have my company's quarterly performance metrics for revenue, gross margin, and customer acquisition rate, and I have benchmark data for top 5 competitors.

Your task:

1. Normalise all data for fair comparison (e.g., currency conversion, adjusting for fiscal year differences).
2. Create comparative bar charts showing my company vs each competitor for each KPI (Key Performance Indicator).
3. Calculate % variance from industry average for each KPI.
4. Identify areas where my company is above average and where it's lagging.
5. Provide 5 targeted recommendations to close performance gaps.

A SECTION 8 – NON-PROFIT ORGANIZATION REGISTERED UNDER THE COMPANIES ACT, 2013

**Output format:** A 2-page PDF competitive report with visual comparisons, an executive summary, and a prioritised action list."

#### Prompt 15 — Data Pipeline Performance Optimisation Plan

"You are a **Data Engineer** experienced in optimising ETL (Extract, Transform, Load) pipelines for speed and efficiency. I have a nightly pipeline that ingests sales, inventory, and customer data into a central warehouse.

Your task:

1. Profile the current pipeline to identify slow queries, inefficient joins, and bottleneck processes.
2. Recommend improvements in query optimisation, indexing, and caching.
3. Suggest parallelisation or batch processing strategies to reduce runtime.
4. Propose monitoring tools to track pipeline health and error rates.
5. Provide an example optimised SQL query and ETL script snippet.

**Output format:** A technical optimisation plan with a "before vs after" runtime projection chart and sample code."

#### Prompt 16 — Social Media Engagement Analytics Dashboard

"You are a **Digital Analytics Expert** skilled in API integrations and BI dashboarding. I have social media engagement data from Facebook, Instagram, and LinkedIn for the past 12 months.

Your task:

1. Create a unified dashboard showing platform-wise engagement metrics (likes, comments, shares, saves).
2. Add a filter to view engagement by post type (video, carousel, single image, story).



3. Highlight top 10 performing posts across all platforms with engagement breakdown.
4. Add follower growth trend lines for each platform.
5. Include an insights section suggesting which content format drives the highest engagement.

**Output format:** Dashboard wireframe + API integration guide + engagement insights report."

### Prompt 17 — Real Estate Market Analysis with Investment Insights

"You are a **Real Estate Data Analyst** with expertise in property market trends. I have a dataset of property sales in [city] for the last 5 years with columns for location, property type, size, sale price, and date.

Your task:

1. Analyse price trends by property type (apartment, villa, plot).
2. Map high-growth neighbourhoods using price appreciation over time.
3. Identify seasonality patterns in sales volume.
4. Calculate ROI projections for top 5 emerging areas.
5. Provide investment recommendations for buyers targeting high rental yield vs capital appreciation.

**Output format:** A PDF market report with heatmaps, trend charts, and a 1-page "Investor Recommendations" summary."

### Prompt 18 — Healthcare Operational Efficiency Analysis

"You are a **Healthcare Data Analyst** working on hospital efficiency improvement. I have anonymised patient visit data, bed occupancy records, and treatment timelines for the past 2 years.

Your task:

1. Calculate average patient wait time, treatment time, and discharge time.
2. Create bed occupancy rate visualisations by department.
3. Identify peak patient inflow periods and staffing shortages.
4. Suggest scheduling optimisations to reduce bottlenecks.
5. Recommend operational changes to improve patient throughput without compromising care quality.

**Output format:** A dashboard layout plan + operational improvement report."

### Prompt 19 — Sales Funnel Drop-off Analysis

"You are a **Marketing Data Analyst** focused on conversion rate optimisation. I have e-commerce funnel data for the last quarter showing visits, product views, add-to-cart events, checkout starts, and purchases.

Your task:

1. Calculate conversion rates for each funnel stage.
2. Identify the stage with the highest drop-off rate and quantify the loss in potential revenue.
3. Analyse patterns in drop-off by device type, browser, and traffic source.
4. Suggest at least 5 tactics to improve conversions at the weakest stage.
5. Provide a visual funnel chart showing current vs projected performance if improvements are implemented.

**Output format:** A funnel analysis dashboard + improvement recommendation document."

### Prompt 20 — Education Performance Dashboard for Institutions

"You are an **Education Analytics Specialist**. I have school-level student performance data for grades, attendance, and extracurricular participation across multiple branches.

Your task:

1. Create visualisations showing average performance by subject and grade level.
2. Add attendance heatmaps highlighting periods of low attendance.
3. Identify correlations between extracurricular participation and academic performance.
4. Highlight top 5 branches in overall performance and bottom 5 for improvement focus.
5. Recommend targeted interventions for low-performing schools.

**Output format:** Dashboard wireframe + insights brief for school management."

### Prompt 21 — Energy Consumption Pattern Analysis for Cost Saving

"You are an **Energy Data Analyst**. I have hourly electricity consumption data for a manufacturing facility over 24 months.

Your task:

1. Identify peak and off-peak consumption periods.
2. Analyse seasonal patterns in energy usage.
3. Quantify potential savings from shifting operations to off-peak hours.
4. Suggest renewable energy integration opportunities.
5. Create a projection model for energy cost savings over the next 12 months.

**Output format:** PDF energy audit report + visual trend charts + savings projection table."

### Prompt 22 — Market Basket Analysis for Cross-Selling

"You are a **Retail Data Mining Specialist**. I have point-of-sale transaction data with item-level details for the past 12 months.

Your task:

1. Use association rule mining (Apriori or FP-Growth) to identify frequent item combinations.
2. Calculate support, confidence, and lift for each rule.
3. Highlight top 10 product pairs with highest cross-sell potential.
4. Suggest bundle offers based on analysis.
5. Project potential revenue increase from implementing top 3 bundles.

**Output format:** Association rules table + actionable cross-sell strategy document."

### Prompt 23 — Website Traffic and Conversion Analytics

"You are a **Web Analytics Consultant**. I have Google Analytics data for my e-commerce site over the past 6 months.

Your task:

1. Identify top 5 traffic sources and their respective conversion rates.
2. Analyse bounce rate, average session duration, and pages per session.

3. Map the customer journey from landing page to purchase.
4. Highlight underperforming landing pages and suggest optimisation strategies.
5. Provide projected improvement metrics if changes are implemented.

**Output format:** Data Studio dashboard layout + optimisation recommendations report."

### Prompt 24 — Manufacturing Process Efficiency Visualisation

"You are a **Manufacturing Data Engineer**. I have IoT sensor data for multiple machines in a production line over the past year.

Your task:

1. Visualise machine uptime/downtime as a Gantt chart.
2. Identify bottlenecks in production flow.
3. Calculate defect rates per machine and per shift.
4. Recommend preventive maintenance schedules.
5. Suggest workflow changes to increase throughput without adding resources.

**Output format:** Factory floor dashboard layout + process improvement plan."

### Prompt 25 — Financial Performance Storytelling for Investors

"You are a **Financial Data Storyteller**. I have quarterly income statements, balance sheets, and cash flow statements for the past 3 years.

Your task:

1. Visualise revenue, gross profit, and net profit trends over time.
2. Calculate key financial ratios (ROE, ROA, current ratio, debt-to-equity) and explain their meaning.
3. Highlight major changes in expenses or revenue sources.
4. Provide a year-over-year growth summary.
5. Frame findings in a narrative that inspires investor confidence.

**Output format:** Investor-ready slide deck with visuals, ratio analysis, and growth narrative."