🚀 Data Project Idea Bank – By Role (With Tools, Steps & Use-Cases)

This curated project idea bank is designed to help aspiring Data Analysts, Data Engineers, and Business Analysts build solid portfolios with practical, business-driven, and interview-ready projects.

# 🔷 Data Analyst Projects

## 📌 Sales Performance Dashboard (Beginner)

🔸 Use Case: Monitor monthly sales, category trends, and target achievements using interactive filters.

🔧 Tools: Power BI, SQL, Excel

🧩 Suggested Steps:

* • Collect transactional sales data from CSV or a mock API.
* • Clean and preprocess using Excel or SQL.
* • Create Power BI visuals: KPIs, clustered bar chart, and region filters.
* • Add slicers for quarter and category segmentation.
* • Publish dashboard and explain KPIs in a LinkedIn post.

## 📌 HR Attrition Tracker (Intermediate)

🔸 Use Case: Identify high-churn departments and factors causing resignation using historical data.

🔧 Tools: Tableau, Pandas, Excel

🧩 Suggested Steps:

* • Import HR dataset and clean nulls, categorize tenure buckets.
* • Use Tableau to build charts showing attrition by department, gender, satisfaction level.
* • Add filters for education level, job role, and marital status.
* • Highlight key insights and HR recommendations.

# 🔷 Data Engineer Projects

## 📌 API to Delta Lake ETL Pipeline (Intermediate)

🔸 Use Case: Automate data extraction from a free API, transform it, and store in Delta format on Databricks.

🔧 Tools: Python, PySpark, Databricks, Delta Lake

🧩 Suggested Steps:

* • Use Python to fetch data via `requests` or a public API (e.g., football-data, crypto, air quality).
* • Clean/normalize JSON into tabular format using PySpark.
* • Write to Delta table in Databricks.
* • Use AutoLoader or schedule batch jobs via notebook or workflows.
* • Optimize storage with ZORDER or partitioning.

## 📌 Real-Time Stream Processor (Advanced)

🔸 Use Case: Ingest and process streaming events like tweets or sensor data using Spark Streaming.

🔧 Tools: Kafka, Spark Structured Streaming, Delta Live Tables

🧩 Suggested Steps:

* • Create Kafka producer to stream JSON messages (mock IoT or logs).
* • Use Spark Structured Streaming to read from Kafka and apply transformations.
* • Sink data into a Delta Lake bronze-silver-gold layer model.
* • Monitor job using Databricks dashboard.

# 🔷 Business Analyst Projects

## 📌 Product Recommendation Analysis (Intermediate)

🔸 Use Case: Analyze cart behavior and co-purchase trends to build product bundles.

🔧 Tools: SQL, Excel, Power BI

🧩 Suggested Steps:

* • Use sales/order\_items data to identify co-occurring product\_ids.
* • Create a matrix of products frequently bought together.
* • Use Power BI to show product affinity with visual heatmaps.
* • Create business suggestions for bundling or cross-selling.

## 📌 Churn Insights & Retention Planning (Intermediate)

🔸 Use Case: Use customer usage trends to create churn buckets and retention alerts.

🔧 Tools: SQL, Excel, Tableau

🧩 Suggested Steps:

* • Define churn logic (e.g., no activity in 60 days).
* • Create churn vs active customer breakdown by segment.
* • Use Tableau dashboard to visualize churn rate trend over time.
* • Highlight potential retention strategies by age group or plan type.

# 🔹 Additional Data Analyst Projects

## 📌 Marketing Campaign ROI Analysis (Intermediate)

🔸 Use Case: Analyze the performance of different marketing campaigns and compute ROI for each channel.

🔧 Tools: Excel, SQL, Power BI

🧩 Suggested Steps:

* • Collect marketing spend and revenue data across different channels.
* • Calculate ROI = (Revenue - Cost) / Cost.
* • Segment by platform (email, social media, Google Ads).
* • Create dashboard comparing performance KPIs.
* • Make recommendations on budget reallocation.

## 📌 Survey Sentiment Dashboard (Advanced)

🔸 Use Case: Analyze customer survey data to identify common sentiment and top complaints.

🔧 Tools: Python, NLTK, Tableau

🧩 Suggested Steps:

* • Clean open-ended survey responses using regex and tokenization.
* • Use sentiment analysis (TextBlob or NLTK).
* • Categorize responses by theme using keyword mapping.
* • Visualize in Tableau with sentiment trends by product/service.

# 🔹 Additional Data Engineer Projects

## 📌 Data Lakehouse Architecture Implementation (Advanced)

🔸 Use Case: Set up a lakehouse model with raw (bronze), cleaned (silver), and aggregated (gold) tables.

🔧 Tools: Databricks, Delta Lake, Python, Airflow

🧩 Suggested Steps:

* • Ingest data from API into Bronze Delta table.
* • Clean and validate in Silver layer.
* • Aggregate KPIs in Gold table.
* • Use Airflow to orchestrate ETL movement across layers.
* • Document pipeline lineage using Databricks or openLineage.

## 📌 ETL Monitoring & Logging System (Intermediate)

🔸 Use Case: Build a system to track ETL job status, errors, and runtime metrics.

🔧 Tools: Python, Airflow, PostgreSQL, Logging module

🧩 Suggested Steps:

* • Use try-except blocks to capture errors during ETL runs.
* • Log job status (Success, Fail) with timestamp into PostgreSQL.
* • Create monitoring dashboard from logs.
* • Send alert emails on failures using SMTP.

# 🔹 Additional Business Analyst Projects

## 📌 Customer Journey Mapping (Intermediate)

🔸 Use Case: Map the end-to-end journey of a user from first visit to conversion using behavioral data.

🔧 Tools: Excel, SQL, Miro/FigJam, Power BI

🧩 Suggested Steps:

* • Extract customer-level events from product logs.
* • Group actions by stage: Visit > Sign Up > Add to Cart > Purchase.
* • Build visual map in Miro or FigJam.
* • Analyze drop-offs and improvement areas using Power BI.

## 📌 Subscription Plan Upgrade Analysis (Intermediate)

🔸 Use Case: Evaluate which features or events lead users to upgrade plans.

🔧 Tools: SQL, Excel, Tableau

🧩 Suggested Steps:

* • Track user activity leading up to upgrade date.
* • Use cohorts to compare upgrade rates based on features used.
* • Highlight trends like ‘3 logins/week leads to upgrade’.
* • Present strategic feature insights in a dashboard.