**🟢 Beginner–Intermediate Projects**

1️⃣ **🎧 Spotify Playlist Dynamics Tracker**

**Data Engineering:** Stream play event webhooks via Kafka, batch process counts in Airflow, and snapshot trends in Snowflake.

**Data Analysis:** Use Python (Pandas, Seaborn) to visualize playlist growth trends, listener retention curves, and track popularity ranking changes over time. Create a Power BI dashboard to show most-streamed songs, peak listening hours, and trending genres.

2️⃣ **✈️ Flight Price History Archiver**

**Data Engineering:** Poll Skyscanner/Google Flights APIs daily via Airflow, structure itinerary data, and build historical dataset in Snowflake.

**Plan to do it:** I will set up MWAA in AWS and it will fetch data using API and store it in S3. File upload to s3 will trigger lambda function that will append the data into snowflake table.

**Data Analysis:** In Power BI, build interactive dashboards showing seasonal price trends, cheapest destinations by month, and volatility heatmaps. Use Python to detect outlier price drops for alerting.

3️⃣ **🎮 Multi-Platform Game Release Tracker**

**Data Engineering:** Scrape Steam/Epic/Xbox store pages with Python, detect new releases via change data capture, and sync to Snowflake.

**Data Analysis:** Use Python (Matplotlib, WordCloud) to analyze release frequency by platform, most common genres, and average price trends over time. Power BI for comparative genre/platform analytics.

4️⃣ **🗃️ AWS S3 to Redshift Incremental Load Orchestrator**

**Data Engineering:** Detect new S3 files, transform with Python in Docker tasks, and load deltas into Redshift via Airflow.

**Data Analysis:** Build Redshift SQL queries to analyze data growth trends, incremental data quality, and freshness metrics. Create Power BI reports to show ingestion performance and file arrival patterns.

5️⃣ **📊 YouTube Metadata Pipeline**

**Data Engineering:** Use Airflow to pull YouTube API JSON, flatten nested structures into a star schema via Python & AWS Glue, and load incrementally into Snowflake.

**Data Analysis:** Use Power BI to analyze video engagement metrics, trending tags, and category-wise performance. Python (Plotly) for visualizing subscriber growth vs. upload frequency correlations.

6️⃣ **💬 On-Demand YouTube Comments Archiver**

**Data Engineering:** Fetch comments via Python, queue in Kafka, batch ingest into Snowflake, and schedule runs with Airflow.

**Data Analysis:** Apply sentiment analysis in Python (NLTK, TextBlob) to classify comments. Use Power BI to display sentiment trends over time and most-discussed topics.

7️⃣ **🌦️ Global Weather Data Hub**

**Data Engineering:** Ingest NOAA/OpenWeatherMap APIs with NiFi, unify schemas across sources, and load to Snowflake for climate trend analysis.

**Data Analysis:** Use Python (Matplotlib, Seaborn) to visualize long-term climate trends and anomalies. Build Power BI dashboards showing temperature extremes, rainfall patterns, and heatwave frequency.

**🟡 Intermediate Projects**

8️⃣ **🛍️ Geo-Fenced Retail Analytics Dashboard**

**Data Engineering:** Enrich purchase logs with geodata in NiFi, stream through Kafka, and aggregate in Snowflake for BI.

**Data Analysis:** Build Power BI maps showing purchase density by location and time. Use Python clustering (scikit-learn) to detect customer hotspot patterns.

9️⃣ **🔌 EV Charging Station Availability System**

**Data Engineering:** Poll charging network APIs with Airflow, stream live status to Kafka, and expose real-time SQL views in Snowflake for route planners.

**Data Analysis:** Use Power BI for real-time station utilization dashboards. Python for forecasting usage peaks based on historical trends.

🔟 **💳 Credit Card Transaction Anomaly Feeder**

**Data Engineering:** Consume Visa/Mastercard event streams via Kafka, flag high-risk transactions using SQL rules, and load to Snowflake for auditor dashboards.

**Data Analysis:** Apply Python anomaly detection models (Isolation Forest) for fraud risk scoring. Power BI to display fraud cases by region, merchant type, and transaction time.

1️⃣1️⃣ **🔐 Real-Time GDPR Compliance Pipeline**

**Data Engineering:** Ingest user data streams from Kafka, auto-redact PII with AWS Lambda/Python, and load sanitized records into Snowflake with audit logs.

**Data Analysis:** Build compliance reports in Power BI showing redaction frequency, data source patterns, and processing latency metrics.

1️⃣2️⃣ **📈 Crypto Exchange Order Book Synchronizer**

**Data Engineering:** Stream order book data from Binance/Kraken APIs via Kafka, normalize bid/ask spreads with Python, and maintain real-time tables in Snowflake.

**Data Analysis:** Use Python (Plotly) to visualize bid/ask spread changes and volatility. Power BI for historical trend comparison between exchanges.

1️⃣3️⃣ **🕹️ Real-Time e-Sports Match Telemetry Pipeline**

**Data Engineering:** Ingest game event streams via Kafka, transform with NiFi, and store/query high-velocity data in Snowflake.

**Data Analysis:** Build Python visualizations of player performance metrics over time. Power BI to show win-rate trends by team and event frequency heatmaps.

1️⃣4️⃣ **🏃‍♂️ Real-Time Player Fatigue Model**

**Data Engineering:** Stream match stats via Kafka, compute fatigue indicators in Python, and store rolling metrics in Snowflake for coaches.

**Data Analysis:** Python (Seaborn) to visualize fatigue progression curves. Power BI dashboards for comparative player health metrics across matches.

1️⃣5️⃣ **🚨 Dynamic Ad Click Fraud Detector**

**Data Engineering:** Stream click events through Kafka, run Python UDF anomaly detection, and archive flagged records in Snowflake.

**Data Analysis:** Python to evaluate detection accuracy with precision/recall metrics. Power BI to visualize fraudulent click patterns by campaign and region.

1️⃣6️⃣ **⚙️ Serverless ML Model Retrainer**

**Data Engineering:** Schedule Airflow DAGs to trigger AWS Lambda retrains on fresh Snowflake data and version models in S3.

**Data Analysis:** Power BI to monitor model accuracy trends and retraining frequency. Python to compare performance metrics between versions.

1️⃣7️⃣ **🛫 Real-Time Flight Delay Predictor**

**Data Engineering:** Ingest airline status APIs through NiFi, feature engineer in Python, serve live delay forecasts, and store scores in Snowflake.

**Data Analysis:** Python (Matplotlib) to compare predicted vs. actual delays. Power BI for airport-level delay trend dashboards.

**🔴 Advanced Projects**

1️⃣8️⃣ **🎲 Real-Time Sports Betting Odds Analyzer**

**Data Engineering:** Ingest live odds from bookmaker APIs via Kafka, process streaks/arbitrage with PySpark, and store in Snowflake for dynamic dashboards.

**Data Analysis:** Power BI to display live odds shifts and arbitrage opportunities. Python to backtest betting strategies.

1️⃣9️⃣ **🎤 Live Concert Demand Forecasting**

**Data Engineering:** Stream Ticketmaster event data through Kafka, enrich with Spotify API popularity metrics, and train Snowflake ML models to predict ticket demand spikes.

**Data Analysis:** Power BI dashboards showing predicted demand curves vs. ticket sales. Python for feature importance analysis in demand prediction models.

2️⃣0️⃣ **🧊 Real-Time NFT Trading Monitor**

**Data Engineering:** Consume blockchain event streams with Kafka, transform in Python, schedule cleanup DAGs, and store in Snowflake.

**Data Analysis:** Python (Plotly) for NFT price fluctuation visualizations. Power BI for top-traded NFT collections and wallet activity tracking.

2️⃣1️⃣ **⚡ Real-Time Energy Market Arbitrage**

**Data Engineering:** Ingest live market prices via Kafka, compute spreads in Python, orchestrate trades in Airflow, and log to Snowflake.

**Data Analysis:** Python to model spread distribution over time. Power BI for visualizing high-profit trading opportunities.

2️⃣2️⃣ **⛓️ Block-to-SQL Blockchain Indexer**

**Data Engineering:** Parse Ethereum blocks to relational tables using Python, handle chain reorganizations, and sync to Snowflake via S3 stages.

**Data Analysis:** Power BI to analyze transaction volume trends and miner activity. Python to identify most active smart contracts.

2️⃣3️⃣ **📰 Real-Time Stock News Correlator**

**Data Engineering:** Ingest SEC filings via Kafka, merge with Bloomberg feeds using NiFi, and store timestamped events in Snowflake for regulatory analysis.

**Data Analysis:** Python (NLP) to detect sentiment shifts in news. Power BI for visualizing correlations between sentiment changes and stock price movements.