**Stepwise Plan for 30-Day YouTube Growth Tracking Project**

**1. Define the Objective**

* Compare the **growth of 100 Pakistani YouTube channels** based on **metadata only**.
* Track metrics:
  + Subscribers
  + Total views
  + Video count
* Goal: Identify **fastest-growing channels**, engagement trends, and overall growth patterns.

**2. Prepare the Channel List**

* Gather **channel names** for the top 100 Pakistani creators.
* Resolve **channel IDs** (unique identifiers) to ensure accurate API queries.
* Store this list in a **structured format** (e.g., spreadsheet, JSON, or Python list).

**3. Set Up YouTube Data API**

* Create a project in **Google Cloud Console**.
* Enable **YouTube Data API v3**.
* Generate an **API key**.
* Understand **quota limits** (typically 10,000 units/day) and plan requests to avoid exceeding them.

**4. Decide the Data Storage Method**

* Two main options:
  1. **Daily CSV snapshots**
     + Save channel metadata every day as YYYYMMDD\_channel\_data.csv.
     + Later merge all CSVs into one table for analysis.
  2. **Database / Data Warehouse**
     + Store metadata in **Snowflake**, **PostgreSQL**, or **MySQL**.
     + Structure the table with columns: channel\_id, snapshot\_date, subscribers, views, video\_count.
* For **realistic ETL practice**, a database or Snowflake is preferred.

**5. Plan the ETL Process**

* **Extract**: Fetch channel metadata (subscribers, views, videos) daily.
* **Transform**: Ensure data types are consistent (numbers for counts, dates for snapshots).
* **Load**: Append new data daily to the storage solution.
* Optional: Add a **logging mechanism** to track errors or failed requests.

**6. Schedule Data Collection**

* Run the extraction daily for **30 consecutive days**.
* Options:
  + **Cron job** or **Airflow DAG** if using a server.
  + Manual run if experimenting locally.
* Ensure each day’s snapshot is **timestamped** for trend analysis.

**7. Data Quality Checks**

* Verify:
  + No missing channel IDs.
  + Subscriber, view, and video counts are numeric and non-negative.
  + Snapshots are consistent across all 100 channels.
* Correct any anomalies immediately to avoid corrupt trends.

**8. Data Analysis Plan**

Once 30 daily snapshots are collected:

1. **Compute Daily Growth**
   * Subscribers gained per day
   * Views gained per day
   * New videos per day
2. **Calculate Growth Rates**
   * Average daily growth over 30 days
   * Percentage growth relative to starting values
3. **Ranking and Trend Detection**
   * Rank channels by fastest-growing subscribers
   * Rank by fastest-growing views
   * Identify patterns (steady growth vs sudden spikes)
4. **Visualize Trends**
   * Line charts: subscriber growth over time
   * Bar charts: cumulative growth comparison
   * Scatter plots: views vs subscribers

**9. Optional Advanced Analysis**

* Engagement ratios: views/subscriber to measure audience interaction.
* Identify channels with high activity but low growth.
* Detect viral spikes in certain channels.

**10. Reporting**

* Create a **summary report** of findings:
  + Top 5 fastest-growing channels
  + Average subscriber growth across all channels
  + Interesting insights (e.g., which niches grow fastest)
* Use Power BI, Tableau, or matplotlib/seaborn in Python for visualization.

**11. Project Review**

* Evaluate the **ETL pipeline efficiency**.
* Check for **missing data or failed API requests**.
* Consider **extending project**: track video-level statistics, engagement, or even comments for deeper insights.