DATA 607 – Data Acquisition and Management

CUNY School of Professional Studies

Project 3 – Part 1: Group Plan

Team Members: Kevin Martin, Joao DeOliveira, Jacob Shapiro, Candace Grant

# Project Motivation

Our project explores which data science skills attract the most interest over time. We will use Google Trends to measure weekly search volume for core skills such as Python, R, SQL, and Tableau to infer demand or popularity levels. We aim to capture both short-term and long-term trends and shifts by examining search frequency across multiple years and regions.

# Collaboration & Tools

We will use Slack for communication, GitHub for scripts and CSVs, and Google Docs for notes and drafts. These tools will help us stay coordinated and transparent throughout the project.

# Data Source & Approach

We are using Google Trends to analyze weekly interest in major data science skills (Python, R, SQL, and Tableau). Data will be exported as CSVs with consistent region and time-window settings, cleaned and reshaped to long format in R, and then stored in a normalized relational database (SQLite). Checks for missing weeks, inconsistent filters, and extreme outliers will be part of the data validation process.

# Roles & Responsibilities

Each member plays a key part in our workflow. Roles can shift as needed, but this initial structure helps guide our collaboration and keep deliverables on track.

| Team Member | Role | Key Responsibilities |
| --- | --- | --- |
| Kevin Martin | Lead / Integrator | Oversees coordination and final assembly. Writes the final summary and prepares the submission files. |
| Joao DeOliveira | Data Cleaning & Formatting Lead | Cleans and reshapes Google Trends CSVs; ensures consistency in column names, regions, and dates. |
| Candace Grant | Visualization & Presentation Support | Builds trend charts and visuals for presentation; supports design and formatting consistency. |
| Jacob Shapiro | Data Validation & Summary Writer | Reviews data for errors or missing values; drafts short summaries of early findings. |

# Logical Model (Normalized Tables)

We will use a normalized logical schema consisting of:

• skills(skill\_id PK, skill\_name)  
• trend\_queries(query\_id PK, skill\_id FK, geography, time\_window, granularity)  
• trend\_data(point\_id PK, query\_id FK, date, interest)

# ER Diagram

See the accompanying file Project3\_ER.pdf for the visual representation of these relationships.

# Timeline & Next Steps

By Sunday: Confirm skills and submit Part 1 (Group Plan + Executive Summary + ER Diagram).  
Next Week: Joao cleans → Candace charts → Jacob summarizes → Kevin integrates → All review for presentation.