Create an API gateway to run SQL queries on the Survey Datasets and retrieve the results in a user-friendly form like JSON

1. Track:
Data Dissemination

2. Description

National Statistics Office, Ministry of Statistics and Programme Implementation (MoSPI) hosts one of the most comprehensive repositories of microdata available at https://microdata.gov.in. This includes anonymized unit-level datasets from flagship surveys like the Periodic Labour Force Survey (PLFS), Household Consumer Expenditure Survey (HCES) conducted by National Sample Survey Office. These datasets are critical for evidence-based policymaking, academic research, and public discourse. However, their accessibility is limited to full-file downloads (CSV, Excel, SPSS, Stata), requiring users to manually filter and process datasets which is a time-consuming process and a barrier to users. Data is available only as full downloadable files and Users must manually filter/analyze datasets locally. No APIs or query-based access exist for developers or researchers. No dynamic visualization, real-time tabulation, or integration with analytical workflows. No mechanism exists for role-based access, rate limits, or tiered usage. Each dataset has a different structure, requiring manual handling.

3. Expected Outcomes/Solutions

Build a scalable, configurable, and privacy-compliant API Gateway that:
- Ingests MoSPI microdata into a structured database
- Offers RESTful APIs for querying filtered subsets of data
- Supports survey-specific configurations for variables, metadata, and relationships
- Provides role-based access with limits based on data volume accessed
- Optionally integrates a micro-payment model for premium access
- Ensures privacy-preserving mechanisms
- Presents results in developer-friendly (JSON/CSV) and visual formats (charts/tables) for end users

4. Relevance to National Priorities or Ongoing MoSPI Initiatives

Aligns with UN Fundamental Principles of Official Statistics and mirrors global efforts like Eurostat SDMX APIs, US Census API, World Bank DataBank. Demonstrates India's leadership in open, controlled, privacy-aware public data access. Supports MoSPI's vision of Statistics-as-a-Service and Digital Public

Infrastructure (DPI) for data.

5. Background Resources or Datasets (if available)

Data sets and Meta data are available on the micro data portal of MoSPI

6. Key Features Required

- Structured Database Ingestion: Load datasets into a relational DB and preserve metadata
- Configurable Survey Framework: Use metadata/config files to define structure, relationships, and filters
- RESTful API Layer
(e.g., api/plfs/data?state=Maharashtra&gender=female&age=15-29)
- Optional Web Interface: Simple UI to create filters and generate tables/charts
- Access Control & Usage Metering: Rate-limiting, volume caps, usage tracking
- Optional Micro-Payment Feature: Simulate pricing model with test gateways
- DPDP-Aligned Privacy Features: Cell suppression, variable aggregation, query blocking
- Developer Experience: OpenAPI/Swagger documentation, Postman collection

7. Bonus Features / Future Scope

Reusable architecture for other government datasets. Opens the door for data-based services and ecosystem growth.

8. Impact Potential

- Time-to-insight reduced
- Equitable access for citizens, researchers, policymakers
- Demonstrates India's capability in scalable data access infrastructure