Milestone-2

Title:

Enhancing Student University Choices with IPEDS Admission Data

Project Idea:

In this project, we will analyze the 2013 IPEDS dataset, which offers comprehensive information about US universities. The dataset covers critical aspects such as the highest degree offered, total applications, admissions, and enrollments for each university. Our primary objective is to empower prospective students to make informed decisions when selecting a university that aligns with their standards and preferences. The key tasks include data extraction, in-depth analysis, data visualization, and providing actionable recommendations to help students make informed decisions when selecting a university based on their preferences.
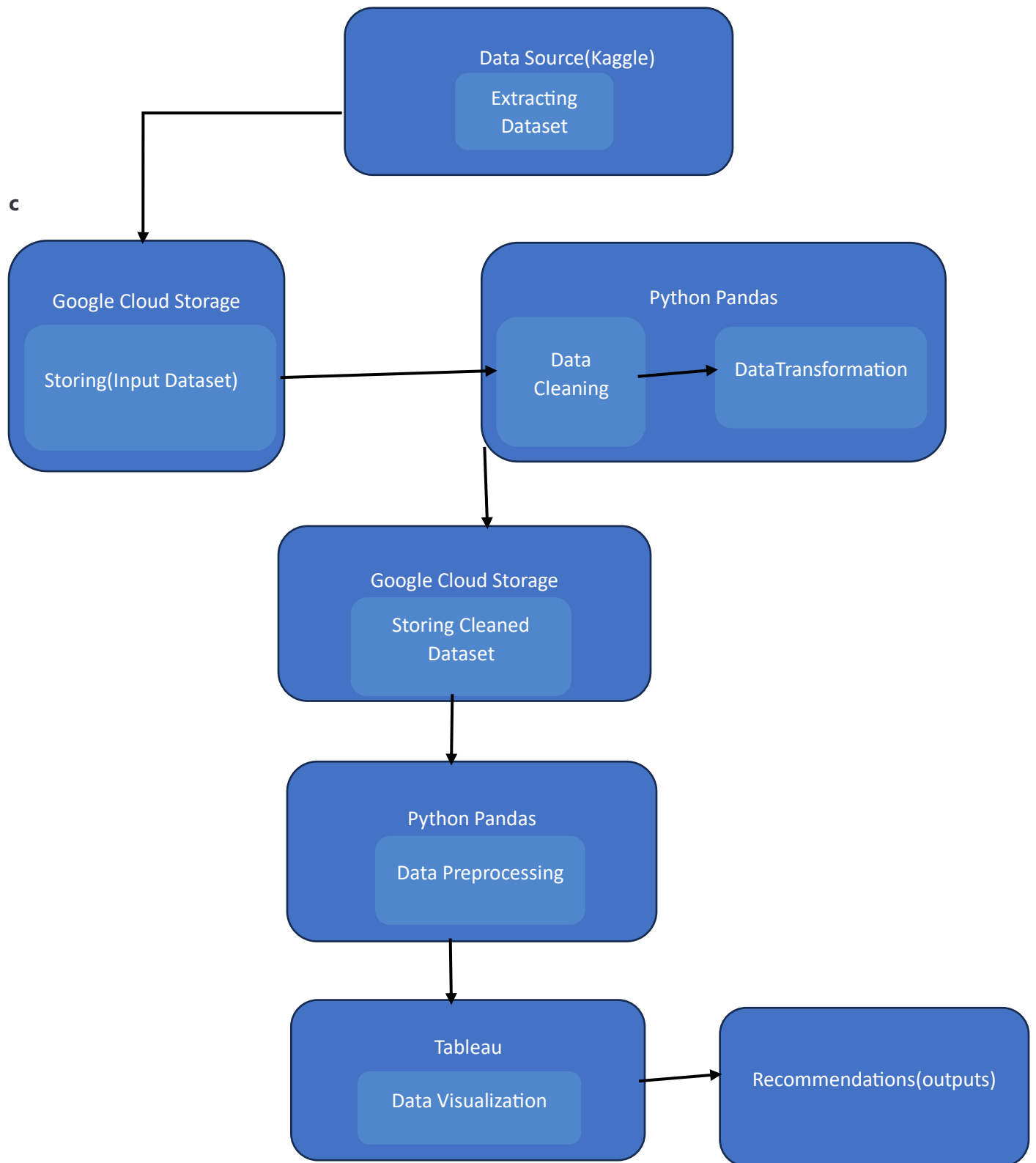
Technology Summary:

Google Cloud Storage for storing Data set.

Python Pandas for Data Cleaning and Transformation and Data Preprocessing.

Tableau for Data Visualization.

Architecture Diagram:

```
┌─────────────────────────────────┐
│      Data Source(Kaggle)         │
│   ┌───────────────────┐          │
│   │    Extracting     │          │
│   │     Dataset       │          │
│   └───────────────────┘          │
└─────────────────────────────────┘

c

┌──────────────────────────┐        ┌─────────────────────────────────────┐
│   Google Cloud Storage    │        │          Python Pandas               │
│                           │        │  ┌──────────┐    ┌────────────────┐ │
│  ┌─────────────────────┐  │        │  │   Data   │───▶│ DataTransformation│
│  │ Storing(Input Dataset)│ │───────▶│ Cleaning │    │                │ │
│  └─────────────────────┘  │        │  └──────────┘    └────────────────┘ │
└──────────────────────────┘        └─────────────────────────────────────┘

                    ┌─────────────────────────────┐
                    │     Google Cloud Storage     │
                    │   ┌─────────────────────┐    │
                    │   │   Storing Cleaned    │    │
                    │   │     Dataset          │    │
                    │   └─────────────────────┘    │
                    └─────────────────────────────┘

                    ┌─────────────────────────────┐
                    │        Python Pandas         │
                    │   ┌─────────────────────┐    │
                    │   │  Data Preprocessing  │    │
                    │   └─────────────────────┘    │
                    └─────────────────────────────┘

        ┌─────────────────────────┐      ┌─────────────────────────────┐
        │        Tableau           │      │  Recommendations(outputs)    │
        │  ┌───────────────────┐   │─────▶│                             │
        │  │ Data Visualization │   │      │                             │
        │  └───────────────────┘   │      │                             │
        └─────────────────────────┘      └─────────────────────────────┘
```

Architecture Summary:

Data Source (Kaggle): We begin by obtaining the 2013 IPEDS dataset from Kaggle, which contains valuable information about US universities.

Data Storage (Google Cloud Storage): We store the original dataset in Google Cloud Storage to ensure it's safely archived and accessible for future use.

Data Cleaning and Transformation (Python Pandas): We use Python Pandas to clean and transform the data. This includes tasks like removing duplicates, handling missing values, and formatting data for analysis.

Store Cleaned Data in GCS (Google Cloud Storage): We archive the cleaned and transformed data in Google Cloud Storage to maintain an organized and pristine copy.

Data Preprocessing (Python Pandas): We perform data preprocessing to prepare the data for effective visualization. This step includes feature selection, data aggregation, and other tasks to make the data visualization-ready.

Data Visualization (Tableau): In this step, we use Tableau to create visual representations of the data, such as charts and graphs.

Recommendations (Output): The insights derived from data visualization are used to generate actionable recommendations for prospective students. These recommendations are our project's primary output.

Goals:

Goal 1: To determine universities with high enrollment.

Goal 2: To determine top colleges with full-time undergraduate enrollment.

Goal 3: To find the total number of universities in each state.

Goal 4: To find the percent of freshmen submitting SAT scores.

Goal 5: To find the top 10 universities with high Full-Time enrollment.