RAJ GUPTA

Python Developer

CONTACT

- 7045093109
- rajguptaa700@gmail.com
- india , mumbai
- www.reallygreatsite.com

PROFILE SUMMARY

As a passionate fresher, I am eager to apply my Python and SQL skills in a real-world setting. I have a strong foundation in writing efficient code, querying databases, and solving data-related problems. I am looking for an internship where I can enhance my technical expertise and contribute effectively to projects

EDUCATION

2024 - 2026
UNIVERSITY OF MUMBAI

- Masters in Data Science

2021 - 2024
UNIVERSITY OF MUMBAI

- Bsc Information Technology

SKILLS

- Python
- Fast API
- SQL
- Git & Giithub
- Docker
- Microsoft Azure
- Pyspark

LANGUAGES

- English: Fluent
- Hindi: Fluent

PROJECTS

Python Weather API
Tech Stacks : - Python , requests , json , wincom

- A weather application in Python fetches real-time weather data using APIs like OpenWeather. It takes user input for a city and displays temperature, Wind Speed in Miles Per Hour (mph)
- This app is built using the requests library to fetch data and I have also added text to speech functionality to this application which makes them more unique . It helps users check the weather easily from their terminal.

PDF Merger
Tech Stacks :- Python , PyPDF2

- A Python merger using PyPDF2 helps combine multiple PDF files into one. It reads each PDF, merges them in order, and saves the final document. The PyPDF2.PdfMerger class makes this process easy by appending PDFs one by one. This tool is useful for combining reports, invoices, or study materials into a single file.

Superstore SQL Project
Tech Stacks :- SQL Server , SSMS , csv

- The SuperStore SQL Server Project is designed to manage and analyze retail sales data efficiently. It involves creating a structured relational database with tables for customers, products, orders, and inventory. SQL queries are used to retrieve sales insights, track inventory levels, and generate reports such as top-selling products and customer purchase trends. The project emphasizes database normalization, efficient indexing, and query optimization for performance. It provides hands-on experience in handling real-world retail data using SQL Server.

Projects

FIle Analyzer Tool using Python :

This Python-based File Analyzer CLI Tool processes text files to provide a summary report. It reads a .txt file, removes punctuation, and filters out common stop words to analyze word frequency. The tool calculates the total word count, top 5 most frequent words, average word length, and total number of sentences. By leveraging re, argparse, and Counter, this script ensures efficient text analysis. It is useful for text processing, linguistic analysis, and data preprocessing in various applications.

Tech Stack :- python, collections, re , argparse

Raj Gupta

Nyc Data Project using microsoft azure cloud :

In this NYC dataset project using Azure Cloud, I leveraged Azure Data Factory (ADF) to dynamically fetch data and store it in Azure Data Lake Storage (ADLS). The raw data was then transformed using Azure Databricks, following the Medallion Architecture. Initially, the data was cleaned and structured in the Bronze layer, then refined and enriched in the Silver layer, and finally optimized in the Gold layer as Delta tables. Once the transformation was complete, I connected the Gold layer to Power BI for visualization, enabling interactive analysis and insights. This architecture ensured scalability, data quality, and efficient processing of the NYC dataset.

Tech Stack :- Azure Databricks , Azure data factory , Azure Data lake gen 2, Azure key vault , Github , delta lakes , pyspark , power bi

Raj Gupta

Apple Analysis Using Pyspark :

The Apple Analysis Using PySpark project focuses on processing and analyzing large datasets related to Apple's product sales, customer reviews. It involves data ingestion from various sources, data cleaning, and transformation using PySpark DataFrame operations. Key insights such as sales trends and customer sentiment analysis are derived through PySpark SQL . The project highlights distributed computing, efficient data processing, and real-time analytics. It provides hands-on experience in handling big data with PySpark in a scalable and optimized manner.

Apple Analysis Using PySpark :- This project gave me hands-on experience in the big data domain. I used the factory pattern to write PySpark code efficiently and gained in-depth knowledge of Delta Tables, from basics to advanced concepts. Additionally, I worked with data extraction from multiple sources, including CSV, Delta Tables, and Parquet, transforming them into a structured format.

Tech Stack :- Pyspark , Databricks Community Edition , Delta Table , Sql

Raj Gupta