Rajdeep Sodha

+91-7208213422 | rajdeep.bsodha444@hotmail.com | Linkedin | Github | Portfolio

EDUCATION

University of Mumbai	Score: 7.83 CGPA
Bachelor of Information Technology	*Aug. 2017 – May. 2020*
MSBTE	Score: 84.76% (Distinction)
Diploma in Information Technology	*Aug. 2014 – June. 2017*

EXPERIENCE

Data Engineer	Aug 2020 – Present
Tata Consultancy Service	*Mumbai, India*

Lloyds Banking Group   *Oct 2022 - Present*

- Leveraged Python, BeautifulSoup, Selenium, Pandas, NumPy, and SQLAlchemy to efficiently scrape, extract, clean, and manipulate data from various UK public registers. Ensured seamless data processing for analysis and storage.
- Expanded data collection capabilities by writing Python scripts to fetch data from external APIs and storing in Oracle DB, enabling access to a wider range of data sources. Enhanced data collection efficiency and enriched the available datasets.
- Implemented simultaneous data cleaning during Python script execution, reducing the over all data cleaning and extraction process from approximately 2 working days to just 5 hours. Increased efficiency and streamlined data preparation.
- Created PowerBI dashboards for the executive team, facilitating data analysis and understanding of data quality. Achieved a significant reduction in error rates by 95%, supporting strategic decision-making at the organizational level.

TCS Ultimatix   *Aug 2020 - Sep 2022*

- Implemented an end-to-end ETL pipeline in Python and SAP HANA,improving customer billing compliance by 80%.
- Optimized ETL script to reduce runtime by 60% and resolve major bottlenecks independently.
- Developed 50+ risk scenarios using Python and SAP HANA, aligning with use cases provided by the finance team.
- Redesigned Jenkins UI to align with user requirements and adhere to TCS standards.
- Executed end-to-end data lifecycle activities with a focus on data gathering, pre-processing, cleaning, analysis, modeling, and visualization.
- Proficient in GCP's BigQuery for efficient data processing and familiar with other big data technologies.

PROJECTS

PyShare | *Python, TKinter*

- Developed a GUI desktop application using Tkinter based on UDP protocol for sharing of files of diverse formats and automatically store them in their respective folders based on the extension type.
- Skillfully utilized os, thread, socket, and subprocess libraries to implement a highly optimized and multi-threaded transfer system, resulting in faster performance.

Worldometer | *Python, Scrapy, SQlite3*

- Developed a Python script leveraging Scrapy to proficiently collect and analyze multi-year population data from various countries.
- Implemented robust data cleaning methologies to ensure utmost accuracy and reliability.
- Effectively stored the processed information in an SQLite database, facilitating seamless data retrieval and enabling informed decision-making.

TECHNICAL SKILLS

Languages: Python, PySpark, Java, SQL (Postgres/Oracle), R
Developer Tools: Git, PowerBI, Docker, Google Cloud Platform, VS Code, PyCharm, Eclipse, Jenkins
Libraries: Pandas, NumPy, Matplotlib, SQLAlchemy, Scrapy, Selenium, BeautifulSoup4

Publication

Bird Species Identifier using Convolutional Neural Network, IJERT, Feb 2021

Certification

Google Cloud Certified - Professional Data Engineer by Google
Qlik Sense Qualification Exam – Business Analyst 2022
MTA: Security Fundamentals - Certified 2017 by Microsoft

Achievement

Awarded star of the month twice at TCS.
Received star team award thrice at TCS.