Abhishek Poojary

Mobile No – (+91) 9819588501
Email ID – abhishekpoojary16@gmail.com

Key Skills & Experience Summary:

- Total Experience 5+ years.
- 3+ Years of Experience as **Data Engineer**
- **Client Communication** & Requirement Gathering.
- Experience in **B2B, BFSI.**
- Expertise in **Data Extraction, Data Cleaning and Data Analysis.**
- Strong knowledge of Programming languages **Python, Pandas and SQL.**
- Experience in **Google BigQuery** and **Data Studio.**
- Strong expertise in **Web scraping** using Python **BeautifulSoup** and **Selenium.**
- Expertise in **Relational Databases** like **MS SQL Server, MySQL.**
- Experience with **NoSQL Database MongoDB.**
- Experience with **AWS stack** – AWS S3 and AWS Lambda.

Career Summary

- A competent professional specialized in building data centric products.
- Skilled in fact collection through discussions to form a holistic understanding about a problem.
- Good industry knowledge, analytical and problem-solving skills and ability to work well within a team as well as an individual.
- Worked in a variety of work cultures, both Corporate MNCs and start-ups. Exposure to global work cultures.
- Skilled in performing data parsing, data cleaning and data preparation.
- Extensive experience in converting unstructured data to structured data.
- Experience with Data Analytics, Data Reporting, Ad-hoc Reporting.
- Familiar with data modelling concepts.

Work Experience

Atyeti Inc. (Client: Factset) (Jan 2021 to Present)
Data Engineer
Responsibilities:
- Working on a ETL pipeline migration project to reduce dependency on the SQL server database.
- Legacy SQL feeds are migrated to Python Pandas based code.
- Identify list of tables and columns list necessary to re-write the db feeds in python.
- Create new Feed Extracts (SQL table dumps) download all necessary raw data and schedule as per demand.
- Use pure pandas, dataframes to mimic SQL operation using RAW input files generated.
- Comparing output data from Migrated feed Vs Prod SSIS package to ensure the ETL pipeline works exactly the same.

Atyeti Inc. (Client: Credit Suisse)(July 2020 to Dec 2020)
Data Engineer
Responsibilities:

- Developed a rich data model for risk factor shocks across all risk asset classes.
- Further enrich the data model with additional metadata, e.g. to better describe countries, regions, currencies, tenors, maturity, credit ratings, and so on.
- Develop spreadsheets to pandas based round trip tools in Python and Pandas.
- Developed a comprehensive ETL pipeline:
    - Extracts data from multiple spreadsheets.
    - Applies mathematical transformations (Incremental to Cumulative, Quarterly to Yearly, Spread to Price based).
    - Validates and loads data to the database.
- Improved existing iterative pandas code by introducing vectorized operations to make transformation significantly faster.
- Created a metadata based unique ID for all risk factors to query data from the database.
- Identify and implement further automation to minimize spreadsheet dependency.
- Write Data workflow documentation.
- Discuss with internal teams and customers for requirement and document gathering. Write business documents to summarize scope of solutions.

Tranzact (September 2019 to February 2020)
Product Engineer
Responsibilities:

- Developed a custom search engine that provides consolidated one-page information on any Indian SME. (Small to medium scale company)
- This tool serves as an information enrichment engine that crawls the web, retrieves information from multiple sources, sanitizes, validates and provides consolidated data.
- Configured google custom search engines.
- Research and compare third party APIs to fetch data.
- Scrape websites using python beautifulsoup library.
- Automate navigation into dynamic websites using Selenium (headless browsers) for web scraping.
- Managed multi-threading, distributed execution for load balancing, effective throttling of scraping operations.

- Data enrichment by extracting useful bits of information from raw text. Performed in Python using regex, string manipulation. Example: Extract founding date from company description.
- Used basic NLP techniques like token frequency and fuzzy string similarity for data validation and analysis.
- Store scraped data into MongoDB and query as needed.
- Define and chase KPIs to improve the engine on a weekly basis.
- Built a data pipeline for the above ETL process.
- Expose the data via an API for consumption by the tech team.
- Liaise with the sales team to understand effectiveness of scraped leads.
- Query data in Google Bigquery and build dashboards in Data Studio to track company's internal data for growth.
- Brainstorm with founders on ideas to grow the network. Studied fundamentals of network effects and read business case studies to contribute to the business side.

Ingram Micro (July 2015 to August 2019)
Responsibilities:
EDI (Electronic Data Interchange) Engineer [July 2015 – Sept 2018]:
- Extract data from unstructured PDF documents, text, emails, spreadsheets and HTML web pages using C#.
- Extract information like document number, address details, order details from purchase order, invoice documents automatically, at scale.
- Load extracted data into MS SQL server database.
- Build web applications in .NET framework to display order information to customers.
- Load validated data into Seeburger Middleware and SAP ERP systems.
- Requirement gathering, preparation of business documents and implementation of EDI (Electronic Data Interchange) setups.
- Generate business reports.
- Liaise with multiple stakeholders to understand specifications of how their system consumes data. Example: Determine mandatory fields in client's purchase order document.
- Write BRD (Business Requirement Document) based on above.

- Translate BRD into Functional documents for tech teams to configure middleware to generate documents with required specification.

Data Analyst [Oct 2018 – Aug 2019]:
- Visualization and analysis of inter-branch stock transfers for better inventory planning and reduction in freight costs using MS-SQL and Python Django framework.
- Reseller propensity modelling to predict which resellers are more likely to purchase a product based on industry segment and product portfolio.
- Use internal and external data sources to fetch additional metadata on customers.
- Use string matching techniques like Levenshtein distance to ensure metadata is mapped against proper company.
- Classify customers into buckets based on industry focus, product offerings, demographics.
- Perform in-depth analysis using SQL, Pandas and Microstrategy dashboards to identify growth drivers in each segment.
- Convey findings to Key business stakeholders using dashboards, infographics.

Technology Expertise:

- Operating Systems: Windows
- Data wrangling Tools: Python, Pandas, Jupyter notebook, SQL.
- Database: MySQL, MS SQL Server, MongoDB
- Domain Expertise: B2B, Indian SMEs, Banking & Finance.

Educational Qualifications:
BACHELOR'S IN COMPUTER ENGINEERING (2015), Fr Conceicao Rodrigues College of Engineering, Mumbai
65 %
INTERMEDIATE (10+2) IN SCIENCE (2011), H.S.C
79.5 %
MATRICULATION (2009), S.S.C
90.15 %

Personal Projects:

Tech articles:
https://in.linkedin.com/in/abhishek-poojary-908392116

Code: https://github.com/abhipoo/

Tech articles: https://medium.com/@abhishek_35553

Data Analysis: https://www.kaggle.com/abhipoo