

MANISH PATIL

Long Beach, CA | (562)- 471-9898 | [My Portfolio Website](#) | [LinkedIn](#) | [GitHub](#) | manishcpatil9@gmail.com

Data Analyst and Data Engineer adept at interpreting, analyzing, and engineering data to extract actionable business insights for informed decision-making. Proficient in both analytics and engineering domains, with a strong portfolio showcasing expertise in the field.

WORK EXPERIENCE:

ENGINEERING TECHNICIAN- FAB LAB CSI³

[August 2022 – December 2023]

- Managed daily operations, maintenance, resolved issues in more than 150 machines, guaranteeing uninterrupted functionality during events.
- Collaborated with over 300 elementary and middle school students from across Los Angeles. Executed fabrication technology projects aimed at promoting advanced knowledge for STEM Education at the CISE lab, particularly dedicated to **underrepresented STEM communities**.
- Facilitated coding sessions to enhance student's **Swift** programming language proficiency within **Apple Community Education Initiative(CEI)**.
- Ensured secure instructional setting as **Apple Teacher** promoting interactive learning, breaking down complex code into comprehensible segments and a supportive atmosphere for students to address coding challenges.

DATA ANALYST- YALI INC.

[September 2020 – March 2021]

- Successfully managed construction of a comprehensive and robust financial domain knowledge corpus using Python and data analytics tools.
- Employed **Beautiful Soup** and **Selenium** for precise **data extraction** on vast knowledge base, contributing **25000 NLP annotations** via **Doccano**
- Implemented data cleaning techniques to ensure integrity, accuracy of the financial dataset, resulting in a **15% improvement** in data accuracy
- Maintained automated data pipelines using Python and **Apache Airflow**, streamlining data ingestion and processing workflows for enhanced efficiency and scalability.
- Effectively leveraged Python libraries for in-depth data analysis, resulting in a **20% reduction** in processing time for financial analytics.
- Played a pivotal role in strategically fortifying NLP models with enriched financial datasets, optimizing automated responses in financial sector.

ACADEMIC PROJECTS:

AZURE OLYMPIC DATA ENGINEERING PIPELINE PROJECT: [GitHub link](#)

[June 2023]

- Developed a comprehensive data pipeline utilizing **Azure Data Factory**, leveraging **ETL** principles to extract data from an API, transform it using Azure Databricks with Apache Spark, and load the transformed data into Azure Data Lake Storage.
- Implemented data ingestion processes by configuring **Azure Data Factory** to extract data from a GitHub repository, utilizing HTTP connectivity, and loading it into **Azure Data Lake Storage** for further processing.
- Utilized **Databricks** to efficiently write, execute **PySpark** code for data transformation tasks, including data type conversion and data cleaning.
- Collaborated on building a comprehensive **data engineering** project, integrating various Azure services including Azure Data Factory, Azure Databricks, and Azure Synapse Analytics, to develop an end-to-end solution for processing and analyzing large-scale datasets.

UBER DATA ENGINEERING PIPELINE AND DASHBOARD CREATION: [GitHub link](#)

[February 2024]

- Developed and implemented a **Data Engineering Pipeline** for Uber trip data using Python, Pandas, SQL, and **Google Cloud Platform** services.
- Used Python for transformation logic to convert raw Uber trip data into a dimensional model structure, incorporating fact-dimension tables
- Demonstrated proficiency in **Data Modeling** concepts including fact tables containing quantitative measures and dimension tables containing descriptive attributes, enhancing data analysis capabilities and facilitating efficient data retrieval.
- Deployed and configured a data pipeline tool, **Mage** on a dedicated virtual machine instance within **Google Compute Engine**, enabling the automated execution of transformation code and facilitating seamless data flow between different stages of the pipeline.
- Created interactive dashboards using Google Data Studio (**Looker Studio**) to visualize insights extracted from processed Uber trip data, incorporating features such as filters, scorecards, and interactive maps to facilitate data exploration and analysis.

CRYPTOLENS- AUTOMATED CRYPTO ANALYSIS: [GitHub link](#)

[November 2023]

- Developed a Python script for real-time crypto data retrieval, utilizing JSON normalization for efficient storage in a pandas DataFrame.
- Implemented timestamp addition and used advanced panda's functions for continuous dataset enrichment, ensuring up-to-date record.
- Incorporated **Extract, Transform, Load (ETL)** principles into the project workflow, ensuring seamless extraction of real-time crypto data, transformation for enriched insights, and efficient loading into a structured pandas DataFrame for comprehensive analysis.
- Demonstrated expertise in Python, leveraging pandas, requests and time for efficient **Data Handling**, API interactions automation.
- Applied Seaborn, Matplotlib for captivating crypto data visualization, employing stacking, index resetting for meaningful representations.
- Implemented streamlined data export CSV and JSON, showcasing proficiency in file management for maintaining datasets with data integrity.

TECHNICAL SKILLS:

- **Languages:** Python and Libraries(Pandas, SciPy, MatPlotLib, Seaborn), SQL(SQL Server, MySQL, PostgreSQL), JavaScript, React, Node.
- **Analytical Tools:** Power BI, Tableau, Microsoft Excel (VLookup, Pivot tables, Conditional Formatting), Jupyter Notebook, PySpark, Microsoft Azure (DataBricks, Azure Data Lake, Azure Data Warehouse), Looker Studio, Big Query, Google Cloud Platform(GCP).
- **Technical Skills:** Git, Mage, Airflow, Kafka, Statistics and Probability, Relational Database, Data Mining, Big Data Analytics, Pattern-Trend Identification, Business Intelligence, Understanding of API's, Django.
- **Certifications(link):** Career Essentials in Generative AI Specialization by Microsoft and LinkedIn, Introduction to SQL, Python for Everybody, Data Analyst Bootcamp, Cisco Data Analytics Essentials.

EDUCATIONAL BACKGROUND:

CALIFORNIA STATE UNIVERSITY, DOMINGUEZ HILLS:

[December 2023]

Master of Science in Computer Science

Relevant Coursework: Advanced Big Data & Cloud Computing, Finite Automata, Advanced AI, Object Oriented Design Analysis, Project Management, Advanced Data Structures and Algorithms, Software Engineering Processes, Software Design and Architecture.

UNIVERSITY OF MUMBAI:

[October 2020]

Bachelor of Engineering in Instrumentation Engineering