|  |  |
| --- | --- |
| **Professional Summary**   * Four years Experienced Data Analyst,Data Enginner with a proven track record of leveraging data-driven insights to drive business growth. * Proficient in statistical analysis, data visualization, and predictive modelling. Skilled in SQL, Python programming languages. * Strong communicator and team player, adept at translating complex data into actionable insights for stakeholders. * Proficiency in Big data technologies such as Hadoop, HDFS, MR/YARN, Sqoop * Proficiency in Linux and shell scripting. | |
| **Skillset**   |  |  | | --- | --- | | **Languages** | Python, SQL | | **Big Data Technology** | HDFS, Flume, Kafka, Sqoop, Spark, Hive | | **Frame works** | Apache Spark, Kafka, Pandas | | **Editors** | PyCharm, Jupyter\_Notebook, VS code | | **Version controls** | GitHub | | **Project Methodologies** | Agile, SDLC, Jira | | **Databases** | Oracle, MySQL,CSV, MangoDB | | **Cloud technologies** | Cloudera, AWS, Microsoft Azure | | **Visualization** | PowerBI, Matplotlib, Seaborn, Plotly | | **Web Technologies** | Flask, HTML | | **Scripts** | Python, Shell Scripting | | **Operating System** | Linux, Windows |   **Professional Experience** |

|  |  |
| --- | --- |
| **Company** | **Duration** |
| Maveric Technology | October 2023 - Present |
| Wipro Technology | January 2017 - Nov 2020 |

|  |
| --- |
| **Education** |

|  |  |
| --- | --- |
| **Course** | **University** |
| Bachelor of Computer Application | D. Y . Patil college Pimpri  Affiliated to Pune University |
| Class XII | Mahalsakant College, Akurdi, Pune |
| Class X | Kanya Vidyalaya, Pimpri, Pune |

**Certifications**

* Pursuing: AWS Certified data engineer
* Udemy: Spark, Linux, MySQL, Agile and Jira, Python, Introduction to Big data

**Project Summary**

|  |  |
| --- | --- |
| **Project 1:** Payments Fraud Detection | **Description:**  Analysing the data extensively, on multiple levels and aggregations for Building a sample dataset containing similar distribution of Non Fraudulent transactions and Fraudulent transactions. This rectifies the class-imbalance and helps machine learning model to perform better predictions and better prevent the loss of funds (profit) to fraudsters.  **Roles & Responsibilities:**   * Work on a bunch of data analysis to find scope for improvement. * Co-ordinate with global teams to understand their requirements, understanding the problem working alongside them. * Collaborate with programmers, engineers, and management heads to identify process improvement opportunities, propose system modifications, and devise data governance strategies. * Focusing on continuous improvement on data processing efficiency resource utilization. * Communicate findings and obstacles to team to achieve best approach.   **Tools Used**: Python Packages/Jupyter Notebooks: Python programming used for Data Analysis and Data Visualization Libraries: NumPy, Pandas, Seaborn, Matplotlib, Plotly.  Languages: Python, SQL |
| **Project 2: Customer Segmentation** | **Description:**  It is the practice of dividing a company's customers into distinct groups based on shared characteristics such as demographics, behaviors, needs,  or preferences. The primary goal is to identify high-value segments, understand their needs, and tailor marketing strategies to better address those needs,  thereby enhancing customer satisfaction, loyalty, and business profitability.  **Roles & Responsibilities:**   * Extract data from primary and secondary sources * Work on Data exploration removing corrupted data and fixing coding errors and related problems * Filter Data by reviewing reports and performance indicators to identify and correct code problems * Using statistical tools to identify, analyse, and interpret patterns and trends in complex data sets could be helpful for the diagnosis and prediction   **Tools Used**: Python Packages/Jupyter Notebooks: Python programming used for Data Analysis and Data Visualization Libraries: NumPy, Pandas, Seaborn, Matplotlib, Plotly.  Languages: Python, SQL |
| **Project 3: Data Engineering- Build Data Pipeline** | **Description**:  Objective was to build a Data Pipeline with Ingestion, Data Transformation, Storage and Data analysis pipeline and to test the building a test automation framework solution.  Developed an end-to-end data pipeline in PySpark for CSV data processing, covering ingestion, transformation, storage, and analysis.  Orchestrated tasks using wrapper and watcher scripts, ensured data quality through preprocessing and cleaning, and configured PySpark for seamless interaction with Hive.  Visualized insights in Tableau/PowerBI. Prioritized high-quality data releases for user feedback and acceptance, ensuring rigorous quality standards at each pipeline iteration. Implemented automated testing for thorough integration testing.  **Roles & Responsibilities**:   * Led development of an end-to-end data pipeline in PySpark for CSV data processing. * Performed extensive EDA to study and analyse data and formed cleaning and transformation rules. * Orchestrated tasks using wrapper and watcher scripts. * Ensured data quality through preprocessing and cleaning scripts. * Developed transformation scripts for analysis-ready data. * Configured PySpark for seamless interaction with Hive. * Created Data Warehouse and Datamart. * Visualized insights in Tableau for effective decision-making.     **Tools Used**: Hadoop, Spark framework with programming language python/Scala, Hive as data warehouse, NoSQL DB or MySQL DB, HDFS, vs-code, Jupyter notebook etc. |