

Yukthika Muppaneni

Data Analyst



+1(513)991-5537

yukmuppaneni07@gmail.com

Highly skilled Data Analyst with **3+ years** of experience in data mining, statistical analysis, and data visualization. Adept at leveraging advanced analytics and programming tools such as SQL, Python, and Tableau to transform raw data into actionable insights that drive business decisions.

PROFILE SUMMARY:

- Data Analyst with **3+ years** of experience in extracting, analysing, and transforming complex datasets into actionable insights.
- Proficient in **Advanced Excel**, **SQL queries**, and **Python/R** (with libraries: **Pandas**, **NumPy**, **Matplotlib**, **Seaborn**, **Scipy**) for data manipulation and analysis.
- Experienced in creating interactive dashboards and data visualizations using **Tableau**, **Power BI** (Certified), and **Looker**.
- Skilled in big data technologies such as **Hadoop**, **Apache Spark**, and **Kafka** for handling large datasets.
- Expertise in cloud platforms including **AWS (S3, Redshift)**, **Google Big Query**, and **Azure Data Lake** for scalable data storage and analysis.
- Strong background in **ETL tools: Informatica, Alteryx**, and custom-built **ETL pipelines** for seamless data extraction, transformation, and loading.
- Adept in **Snowflake** for efficient data warehousing and high-performance analytics.
- Specialized in **Machine Learning**, building predictive models and algorithms for insights and improved business outcomes using **scikit-learn**, **TensorFlow**, and **Kera's**.
- Experienced in **PowerShell** and **Bash scripting** for automating tasks and improving data workflow efficiency.
- Hands-on experience with **Apache Airflow** for orchestrating complex workflows and managing data pipelines.
- Familiar with **Docker** and **Kubernetes** for containerization and deployment of data analytics applications.
- Proficient in **Git** for version control, collaboration, and maintaining code integrity across teams
- Solid experience with **Microsoft Azure** services like **Azure SQL Database**, **Azure Synapse Analytics**, and **Azure Machine Learning**.
- Familiar with **Google Cloud Platform (GCP)** tools like **Google Cloud Storage**, **Google Big Query**, and **Dataflow** for cloud-based analytics and data management.
- Experience with **NoSQL databases** like **MongoDB** and **Cassandra** for unstructured data storage and analysis.
- Skilled in data integration tools like **Talend** and **MuleSoft** for connecting and integrating data from multiple sources.
- Proficient in using **Power BI** to design and implement interactive, real-time dashboards and reports, enabling stakeholders to make data-driven decisions with ease and Proficient in using **Apache Kafka** for real-time data streaming, ensuring seamless data integration across distributed systems and enabling fast, reliable data processing at scale.

EDUCATION: University of

Cincinnati, Masters in
Information Technology, USA
from Aug 2023 – Dec 2024

CERTIFICATION:



TECHNICAL SKILLS:

Programming Languages: SQL,

Python, R, JavaScript

Data Analysis & Statistical

Tools: Excel (Advanced), SAS,
SPSS

Data Visualization: Tableau,

Power BI, Looker, QlikView

ETL Tools: Informatica, Alteryx,
Talend, Apache Nifi

Big Data Technologies: Hadoop,
Spark, Snowflake, Hive
Apache.

Machine Learning: Scikit-learn,
TensorFlow, PyTorch, Kera's

Databases & Data

Warehousing: MySQL,

PostgreSQL, Oracle, Mango DB.

Cloud Platforms: AWS, Google
Big Query, Microsoft Azure

Web Scraping & APIs: Beautiful
Soup, Selenium,
Postman

Automation & Scripting: VBA
(Excel), Power Automate, Bash

Version Control: Git, GitHub

**Collaboration & Project
Management:** JIRA,

Confluence, Microsoft Teams

Data Automation & Workflow

Orchestration: Apache Airflow,
Luigi

Data Cleaning &

Preprocessing: Open Refine,
Trifacta.

Dashboarding & Reporting:
Google Data Studio, Zoho
Analytics

WORK EXPERIENCE:

Nationwide| Data Analyst II **Columbus, Ohio | April 2024 – Present**

Description: Nationwide Mutual Insurance Company offers a comprehensive range of products and services, including Property and Casualty Insurance, Life Insurance, Retirement Solutions, Asset Management, Strategic Investments. Collecting, analysing, and interpreting complex insurance data to identify trends, patterns, and key insights, ultimately providing actionable recommendations to improve business operations and developing dashboards, and performing analyses to identify performance drivers.

Responsibilities:

- Analysed and interpreted complex insurance data sets to provide actionable insights for optimizing claims processing, underwriting, and customer service operations at **Nationwide Mutual Insurance**.
- Developed and maintained **SQL queries** and **data models** for efficient data extraction and analysis, enhancing decision-making in insurance claims management and risk assessment.
- Utilized **Power BI** and **Tableau** to create visually appealing and interactive dashboards and reports, helping stakeholders visualize key insurance metrics, such as claims processing times and policyholder engagement.
- Automated regular reporting processes using **Python** and **SQL**, significantly reducing manual intervention and ensuring timely and accurate data dissemination to cross-functional teams.
- Supported data integration efforts by managing and transforming insurance data using **ETL tools** such as **Informatica** and custom **ETL pipelines**, improving data flow across insurance departments.
- Applied **machine learning** techniques, including classification and regression models using **Python** and **scikit-learn**, to predict insurance claim outcomes, optimize risk assessments, and improve fraud detection processes, enhancing operational efficiency at **Nationwide Mutual Insurance**.
- Optimized and automated data pipelines using **Apache Spark** and **Hadoop** to process large volumes of structured and unstructured insurance data, ensuring efficient performance and scalability.
- Implemented and maintained data warehousing solutions using **Amazon Redshift** and **Snowflake**, providing scalable, reliable data storage for insurance claim metrics and customer data.
- Designed and managed data models and schemas for relational databases in **SQL Server**, **Oracle**, and **MySQL**, improving data access and query performance for claims and underwriting teams.
- Developed and deployed **APIs** for seamless data sharing and integration with internal and external insurance systems, ensuring efficient data exchange across platforms using **REST** and **GraphQL**.
- Built and maintained real-time data streaming pipelines using **Apache Kafka** to process and analyse live insurance claims data, enabling real-time monitoring of claims status and service level agreements.
- Developed and maintained interactive dashboards using **Tableau** and **Power BI**, providing real-time insights into insurance claim processing efficiency and policyholder satisfaction for cross-departmental teams.
- Leveraged **Docker** and **Kubernetes** to containerize and deploy data applications in a scalable and efficient manner for insurance data analysis and reporting, ensuring seamless deployment across environments.
- Implemented **CI/CD pipelines** for automated testing, deployment, and monitoring of data applications, ensuring efficient delivery and operation of insurance data products.

ENVIRONMENT: SQL queries, Power BI, Tableau, Python, ETL tools, Informatica, machine learning, scikit-learn, Apache Spark, Hadoop, Amazon Redshift, Snowflake, Oracle, APIs, REST, GraphQL, Apache Kafka, Docker, Kubernetes, CI/CD pipelines.

Doha Bank | Data Analyst **Mumbai, India | Sept 2022 – Jul 2023**

Description: Doha Bank, is one of largest private commercial banks offers a comprehensive range of financial services to individuals, businesses, and institutions both domestically and international. Led the analysis of software performance data to identify patterns, helping optimize application efficiency and user experience and created and delivered regular reports on software performance, user engagement, and feature adoption, guiding product development strategies.

Responsibilities:

- Analysed complex financial data sets to provide actionable insights and drive decision-making across multiple business domains using **SQL** and **Python** for data extraction and analysis.

- Developed and implemented ETL processes to cleanse, transform, and load financial data into various data storage systems using tools like **SQL Server** and **Azure Data Factory**, ensuring seamless data integration.
- Designed and created interactive dashboards and reports for stakeholders using **Power BI** and **Tableau** to visualize key financial metrics and provide real-time insights into financial performance.
- Utilized **Machine Learning** models to identify trends and forecast future financial outcomes, applying **R** and **Python** libraries like **scikit-learn** and **TensorFlow** to enhance predictive capabilities.
- Extracted and manipulated financial data from different sources such as **SQL** and **NoSQL** databases and cloud platforms like **Azure**, performing data mining and statistical analysis using **R** and **Python**.
- Leveraged **Hadoop** and **Spark** for handling large-scale financial data sets, enabling efficient data processing and analysis for improved decision-making in forecasting and risk assessment and worked with **NoSQL** databases like **MongoDB** and **Cassandra** for storing and querying unstructured financial data, supporting dynamic reporting and analysis.
- Implemented and optimized data warehousing solutions in **Azure Synapse Analytics** and **Google Big Query** for faster reporting and analysis, improving data accessibility and reducing query times for financial reporting.
- Built automated reports and dashboards using **PowerShell** and **Python** scripts to streamline regular reporting processes, increasing operational efficiency for finance teams and Integrated data from external financial sources, including third-party APIs, into internal systems, ensuring seamless data flow and enabling comprehensive financial analysis.
- Utilized **Jupyter Notebooks** and **Azure Notebooks** to document, visualize, and share financial analysis findings with stakeholders, ensuring transparency and collaboration.
- Created and automated data quality checks using **Apache Airflow** to monitor and validate financial data processes, ensuring accuracy and reliability across systems.
- Managed and monitored **CI/CD** pipelines for financial data integration and deployment using **Jenkins** and **GitLab**, optimizing workflow automation and deployment efficiency.
- Conducted data profiling and validation using tools like **Talend** and **Alteryx** to ensure the accuracy and completeness of financial data before analysis, improving the integrity of decision-making processes.
- Developed and deployed real-time data streaming applications using **Apache Kafka**, enabling continuous data flow for financial transactions and improving operational monitoring.
- Integrated and optimized **Snowflake** for cloud-based data warehousing, providing efficient and scalable data solutions for financial reporting and analytics
- Utilized **Docker** and **Kubernetes** for containerizing financial applications, enabling easier deployment and scalability of data-driven solutions.
- Applied **Google Data Studio** to design customized reporting solutions for stakeholders, providing them with easy-to-understand visualizations and metrics to support decision-making.

ENVIRONMENT: SQL, Python, Azure Data Factory, Power BI, Tableau, scikit-learn, TensorFlow, MongoDB, Cassandra, Azure Synapse Analytics, PowerShell, Jupyter Notebooks, Apache Airflow, CI/CD, Jenkin, GitLab, Talend, Alteryx, Apache Kafka, Snowflake, Docker, Kubernetes, Google Data Studio.

PharmEasy | Data Analyst **Mumbai, India | May 2021 – Aug 2022**

Description: PharmEasy is an Indian e-pharmacy company offers a comprehensive range of healthcare services, including the delivery of prescription and over-the-counter medicines, diagnostic tests, and telehealth consultations. Monitored customer feedback and transaction patterns to identify potential issues or opportunities for improvement in the customer experience and worked closely with the product team to help them understand customer preferences and suggest areas for new product development.

Responsibilities:

- Collected, extracted, and integrated data from multiple healthcare systems, customer databases, and external sources using **SQL**, **Python**, and **APIs** to ensure seamless data flow across platforms.
- Performed **data cleansing**, validation, and transformation to ensure data accuracy and consistency using **Python (Pandas, NumPy)** and **SQL**, maintaining high-quality healthcare data.
- Developed and maintained healthcare data models to support decision-making processes, utilizing **R**, **Python**, and **SAS** for accurate predictive insights.

- Designed and generated standard healthcare reports, dashboards, and visualizations using **Tableau**, **Power BI**, and **Excel**, providing stakeholders with actionable insights.
- Managed and optimized healthcare data storage and processing systems using **Amazon Redshift**, **S3**, and **Google Big Query**, ensuring quick access to critical healthcare data.
- Designed and maintained automated **ETL** pipelines with **Apache Airflow**, **Talend**, and **AWS Glue**, streamlining data integration processes for the e-pharmacy platform.
- Performed risk analysis and predictive modelling using statistical techniques and machine learning with **Scikit-learn**, **TensorFlow**, and **Keras**, optimizing customer engagement strategies.
- Assisted in budgeting processes and long-term financial forecasting using financial modelling techniques in **Excel** and **Python**, supporting financial planning for the e-pharmacy business.
- Ensured adherence to healthcare regulatory standards and compliance requirements by automating regulatory reporting processes with **Python** and **BI tools**.
- Integrated disparate healthcare data sources into a cohesive dataset for analysis using **Informatica**, **SSIS**, and **Talend**, facilitating a unified view of patient and sales data.
- Managed large-scale healthcare data processing and analysis by leveraging **Hadoop** for distributed storage and processing, ensuring scalability in handling vast amounts of customer and transaction data.
- Developed and deployed machine learning models for patient segmentation, predictive health outcomes, and inventory forecasting using **Scikit-learn**, **TensorFlow**, and **Python**, enhancing personalized healthcare services.
- Designed and optimized healthcare data warehousing solutions using **Snowflake**, ensuring efficient storage, scalability, and high-performance querying for large healthcare datasets.
- Utilized **Python** tools such as **Pandas**, **NumPy**, and **Matplotlib** to perform advanced healthcare data analysis, automate reporting, and visualize trends for strategic decision-making.
- Leveraged **Looker** to create interactive dashboards and reports, enabling real-time healthcare data analysis and insights for improved service delivery and operational efficiency.

ENVIRONMENT: SQL, Python, APIs, Pandas, NumPy, Tableau, Power BI, Excel, Amazon Redshift, S3, Google Big Query, Apache Airflow, Talend, AWS Glue, Scikit-learn, TensorFlow, Keras, BI tools, Informatica, SSIS, Talend, Hadoop, Snowflake, Looker.