

SARTH PATEL

United States | +1 4479022091 | sarthp870@gmail.com | [Portfolio](#) | [LinkedIn](#)

EDUCATION

University of Illinois Urbana Champaign, Champaign, US

Expected May 2026

Master of Science in Information Management

Courses: Information Consulting, Data Warehousing and Business Intelligence, Data, Models and Information

University of Mumbai, India

Jun 2024

Bachelors of Technology in Computer Engineering

Courses: Advance Database Management Systems, Data Mining, Big Data Analytics, Machine Learning, Artificial Intelligence

PROFESSIONAL EXPERIENCE

Data Scientist | Webomatic Softech Private Limited | India

Jun 2022 – Jan 2023

- Orchestrated financial data analysis on over **15,000+** transaction records, identifying crucial trends in market dynamics, and delivering implementable solutions through **PowerBI** dashboards.
- Conducted detailed ad-hoc financial analysis (EDA) using **SQL** and **Python** to uncover key trends, driving initiatives that increased financial efficiency by reducing costs by **9%** and enhancing profitability through a **20%** boost in customer engagement.
- Leveraged **Jupyter Notebook** for exploratory **data analysis**, using **Matplotlib** and **Seaborn** to create interactive visualizations, enabling clear presentation of trading trends and investment insights to team members.
- Performed hands-on data analysis of large data sets to create customer profiles for over a million customers. Used various data mining techniques such as **clustering** and **decision trees**.
- Designed and deployed a financial **KPI** tracking dashboard in **PowerBI**, streamlining monitoring of financial metrics and reducing analysis time by **40%**, enabling data-driven strategies for improved budget management and cost control.
- Developed and implemented data pipelines that curated data from various sources (such as **CSV**, **XML** & **SQL**) to create a data warehouse
- Applied Hypothesis testing such as **t-tests**, **ANOVA**, **Chi-square** and **Tukey's HSD** to analyze high-yield investment data, leading to a **15%** reduction in financial risk exposure.
- Participated in team of 5, in **A/B** testing, code reviews and **agile** methodology which helped to improve development time by **45%**.

PROJECTS

AdventureWorks Sales Analytics | Tech – Azure, Databricks, PowerBI | [Github](#)

- Integrated data from an **API** using **Azure Data Factory**, creating dynamic pipelines to automate the ingestion process and land data into the Bronze layer in **Azure Data Lake** Gen2.
- Utilized **Databricks** to transform and process data from the Bronze layer to the Silver layer, implementing efficient data transformation workflows using **PySpark** for enhanced data quality and usability.
- Designed and implemented a scalable data warehousing solution with **Azure Synapse Analytics** to serve processed data for advanced analytics and integrated it with **Power BI** for dynamic reporting and visualization.

Youtube Data Analytics | [Github](#)

May 2024 – Jun 2024

- Utilized **AWS CLI** to efficiently upload the YouTube videos dataset from Kaggle into an **S3** bucket, ensuring scalable and secure storage of raw data.
- Implemented a robust **ETL** system using **AWS Glue** and **Lambda** to process **JSON** data through **ETL** processes, leveraging **PySpark** scripts, **Hadoop**, and **Hive** for distributed data processing.
- Efficiently transformed and loaded data into parquet format within **S3** Applied **Spark** for large-scale data transformations, optimizing processing speed and reliability. Utilized **AWS Athena** for interactive querying and **Amazon QuickSight** for building dashboards.

NLP Model for Duplicate Quora Question Pairs | Tech– Python, Machine Learning, NLP, Numpy, Pandas | [Github](#)

- Applied **Natural Language Processing (NLP)** techniques, including word embeddings, POS Tagging, and **fuzzy** string matching, to build a machine learning model for detecting duplicate questions.
- Integrated **Feature Engineering** like token based & length based techniques and achieved a **7.84%** performance boost on test data when used with **Random Forest** classifiers.
- Achieved **79.2%** accuracy using **XGBoost**, significantly improving Quora's ability to streamline content.

T20 Cricket World Analytics | Tech – Python, BeautifulSoup, DataModels, Numpy, Pandas, PowerBI, DAX | [Github](#)

- Scraped player statistics from ESPNcricinfo using **Beautiful Soup**, with 400+ data points, ensuring 100% coverage of key performance metrics for analysis.
- Cleaned and transformed raw data using **Pandas** and **Numpy**, reducing data inconsistencies by 95% for in-depth analysis.
- Designed **4** interactive **Power BI** dashboards to showcase player performance metrics, including run rates, wicket counts, and player efficiency, enabling real-time insights.

SKILLS SUMMARY

CERTIFICATIONS

Complete Machine Learning & Data Science Bootcamp 2023 | [Link](#)

Microsoft Power BI Desktop for Business Intelligence | [Link](#)

SARTH PATEL

United States | +1 4479022091 | sarthp870@gmail.com | [Portfolio](#) | [LinkedIn](#)

Python · C++ · R · Microsoft Excel · SQL · Hadoop · Hive · Java · Power BI · Tableau · Alteryx · Visio · SAS · Postgres · Data Analytics · Statistics · Scikit Learn · Snowflake · Azure · Databricks · Snowflake · PySpark · PowerPoint · AWS · Project Management

CERTIFICATIONS

Complete Machine Learning & Data Science Bootcamp 2023 | [Link](#)

Microsoft Power BI Desktop for Business Intelligence | [Link](#)