

Milan Prajapati

Canada | (778)855-1756 | mbprajapati1732@gmail.com | [linkedin.com/in/itsMilan1999/](https://www.linkedin.com/in/itsMilan1999/) | [Portfolio](#)

SUMMARY: Proficient in Data management, Cloud services, visualization, and Dashboard reporting | Committed to delivering hidden insights from data for business growth and strategy-making | Life Long Learning

EDUCATION

Master of Professional Studies (M.P.S.), Data Analytics | **Northeastern University** (Vancouver, BC)

Bachelor of Engineering (B.E.), Computer Engineering | **Gujarat Technological University** (Ahmedabad, IN)

EXPERIENCE

Nursing Community Assessment Service (NCAS)

Vancouver, BC

Data Analyst

January 2023 – Current

Data Processing

- Efficiently maintain database of over 50,000 patient examinations and executed complex SQL queries on a terabyte-scale dataset.
- Enhanced query processing with window functions, joins, and nested queries, and redesigned the database using a Snowflake schema.
- Implemented Data Clusters for efficient Data access and further performed Data ETL (Extract, transform, and load).
- Developed data pipelines using BigQuery, Python, and Apache Airflow speeding up the data processing by 35%.
- Constructed CRM model with real-time data using cloud streaming services with power BI, Visualizing critical information.

Heart Health Prediction Model

- Utilized Python with PySpark to apply feature selection on patient data, pinpointing key factors affecting heart health.
- Developed a data pipeline that combines on-premises SQL data with JSON data from web input, creating a centralized SQL database.
- Trained a prediction model using Python, with identified key factors to anticipate potential heart health issues in advance.
- Employed AIC and BIC comparisons to detect flaws in pre-existing model, leading to 23% enhancement in accuracy for the new model.
- Deployed the finalized model on the cloud, conducted unit testing, and formulated visualization of performance results to show impact.

Northeastern University

Vancouver, BC

Instructional and Research Assistant

January 2023 – December 2023

JobLink Insights

- Processed 28 GB of AI-generated JSON data, designed a SQL star schema, enhancing data usability, and improving query performance.
- Utilized AWS Kinesis data streaming and developed data pipeline to provide continuous input data for training machine learning model.
- Implemented natural language processing and semantic analysis with Python Hugging Face to enhance job recommendation system.
- Tested and deployed Job Recommendation system, boosting suggestion accuracy by 82% through resume and job description analysis.
- Created a dashboard using Tableau to highlight market requirements and currently available opportunities.

Starlink Dataset Analysis

- Configured cloud policies and transferred over 50TB of data to cloud, and performed data cleaning for optimized cloud data storage.
- Implemented ad-hoc architecture and parallel query processing with SQL resulting in improve data accessibility and query processing.
- Calculated satellite launch-to-functioning ratio and average lifespan using Python resulting in optimizing resource allocation by 32%.
- Performed time series analysis on data to identify improvements in space projects, resource usage and utilization.
- Created a detailed report featuring graphs and data visualization to emphasize key findings using Power BI.

Pandavaz Advertising Agency

Gujarat, IN

Data Scientist

July 2019 – September 2022

Credit Score Analysis and Prediction

- Developed Docker containers and images for project setup across diverse systems, enhancing setup process and achieving time savings.
- Utilized data encryption and Key Management Service to securely store financial data, concealing personally identifiable information.
- Created data pipeline using Data Transfer Stream to seamlessly combine cloud and on-premises data for model training and Data backup.
- Build Credit rating model based on financial data, purchases and spending, predicting 93% relevant factors influencing credit scores.
- Implemented prediction models to determine the creditworthiness of individuals, facilitating more informed lending decisions.

Supply Chain Analysis

- Aggregated demand and supply data from Excel, Web and handwrote notes, and build Centralized database for data storage.
- Implemented role-based data access policies and secured the data pipeline in SQL to enhance data security and integrity.
- Applied time series analysis on collected historical data to predict changes in demand resulting in 78% product availability in the market.
- Created interactive web dashboard using Power BI & Excel to digitally interact with research and findings with visual evidence.
- Led meetings explaining research and fundamentals with stakeholders and delivered a final report with insights, and research evidence.

SKILLS

Technical Skills: Python, R, SQL, JavaScript, AWS, Azure, GCP, Prompt Engineering (ChatGPT, Google Bard)

Frameworks: Scikit-learn, Pandas, NumPy, SpaCy, PySpark, Matplotlib, TensorFlow, BeautifulSoup, ggplot2, Shiny

Database & Visualization: MySQL, Oracle, SQL Server, MongoDB, Microsoft Power BI, Microsoft Excel, Tableau, Microsoft Word

Tools: SCRUM (AGILE), Git, GitHub, Docker, Office 365 Suite, JIRA, Slack, Workday, Notion

COMMUNITY INVOLVEMENT & ADDITIONAL INFORMATION

Certifications: AWS Cloud Practitioner | AWS Data Engineer | Tableau for Data Scientists | IBM Data Science | Salesforce for admins

Achievements: HackerRank Python Gold Badge | HackerRank SQL Gold Badge

Research Paper: Job Market Analysis of San Francisco, delivering recommendations for economic growth and worker requirement in city

Learnings: AWS Data Analyst | Prompt Engineering | Static Analysis using R | Cloud Engineering | load balancing and auto-scaling for cloud