Simarpreet Kaur

kaursimar9464@gmail.com | (902) 880-8691 | simarpreet-kaur.com | LinkedIn | GitHub

EDUCATION

McMaster University - Honours Bachelor of Applied Science, Computer Science — Sept 2022 – Present
St Lawrence College - Computer Networking and Technical Support — Sept 2017 – June 2019

CERTIFICATIONS

- Google Advanced Data Analytics Certificate
- The Nuts and Bolts of Machine Learning | Credential ID: 7HHKEHGUJ75Q
- Regression Analysis: Simplify Complex Data Relationships | Credential ID: WDKX9Y6LT5NL
- Data Science Certification – BrainStation | Certificate ID: 830-947-036
- Data Analysis with Python – IBM
- Getting Started with Python – Google | Credential ID: NF83JSFBDQAX
- Go Beyond the Numbers: Translate Data into Insights – Google | Credential ID: SQRBAWG8QK6Q
- The Power of Statistics – Google | Credential ID: NDGLES25BNUY

PROJECT EXPERIENCE

Titanic Survival Prediction – Kaggle
Predicted passenger survival using machine learning with ~80% accuracy on test data.

- Cleaned and processed 890+ entries, engineered features from text and categorical data
- Trained Logistic Regression and Decision Tree models using Scikit-learn
- Evaluated performance using F1-score and accuracy; submitted results to Kaggle competition

Waze User Data EDA and Executive Summary
Uncovered key behavior patterns across 20,000+ Waze user sessions.

- Cleaned and analyzed user activity data in Python (Pandas, Seaborn)
- Visualized trends in session length, drive frequency, and location clusters
- Delivered a data-driven executive summary using the PACE strategy for stakeholders

Employee Retention Prediction – Salifort Motors
Predicted employee turnover using ML models to inform HR decisions and reduce costs.

- Achieved 95%+ accuracy with ML models (Logistic Regression, Random Forest, XGBoost) predicting employee turnover
- Analyzed 15,000+ HR records to uncover attrition drivers like department, workload, and hours
- Derived insights to support retention strategies and lower recruitment costs

Palmer Penguins Clustering (K-Means)
Segmented 344 penguins into species clusters based on physical attributes.

- Applied K-Means clustering and silhouette analysis to identify optimal cluster count
- Visualized cluster separation using scatter plots and centroid markers
- Delivered insights on species traits and model reliability

TECHNICAL SKILLS

Languages & Tools: Python, SQL, Java, C/C++, Pandas, NumPy, Scikit-learn, XGBoost, Tableau, Jupyter
Data Science: Data Cleaning, EDA, Regression, Classification, Clustering (K-Means), Model Evaluation (F1, Accuracy)
Visualization & Communication: Matplotlib, Seaborn, Tableau Dashboards, Data Storytelling, Executive Summaries