

# Arturo Rey

Vancouver, British Columbia



672-558-0655



arturorey.h@outlook.com



arturoreyh



AReyH

[areyh.github.io/arturo-portfolio/](https://areyh.github.io/arturo-portfolio/)

## Summary

Data Scientist with extensive experience in data modeling using machine learning, business intelligence, and managing the overall data flow, data processing and modeling process. Proven ability to develop business solutions by analyzing and interpreting data to create value. Always eager to learn new bleeding edge technologies.

## Experience

### Pacific Salmon Foudation

April 2024 – Present

*Data Scientist Consultant*

*Vancouver, Canada*

- Designed and deployed an interpretable predictive model to accurately identify tagged salmon species, aiding researchers in their identification efforts.
- Implemented Outmigration models to forecast salmon migration patterns, enabling researchers to better allocate resources and predict migration timings.
- Designed and deployed an efficient data pipeline using Apache Airflow, enhancing process automation and data workflow management.
- Developed a Bayesian CJS model to estimate detection probability at each stage, survival probability between stages, and overall survival probability from tagging release to each detection occasion, providing comprehensive insights into salmon survival dynamics.

### Horbath Technologies

April 2022 – December 2023

*Data Scientist*

*Barranquilla, Colombia*

- Spearheaded the development of the data science department within Gases del Caribe, demonstrating leadership and strategic vision. Mentored and coached junior data scientists, fostering their professional growth and development while ensuring the team's success.
- Designed and deployed over 200 predictive models tailored for natural gas distribution stations, aimed at forecasting daily consumption with a one-day advance notice, utilizing a blend of classical and advanced forecasting methods, and Recurrent Neural Networks to ensure accuracy and reliability of predictions. This contributed to optimizing operational efficiency by providing accurate predictions.
- Built and deployed risk prediction models that assigns a risk level to customers, projected to increase project revenue by 20%.
- Increased customer retention and Diamond customer eligibility through the use of RFM analysis and clustering techniques to recommend proper customer segmentations and improve marketing strategies.

### Termocandelaria

December 2019 – April 2022

*Engineer*

*Cartagena, Colombia*

- Led a project utilizing machine learning to develop a predictive maintenance model, significantly reducing maintenance costs and minimizing equipment downtime, thereby enhancing overall operational efficiency.
- Managed a team responsible for the operation and maintenance of the plant's production equipment, resulting in increased production uptime and reduced downtime, ultimately boosting productivity and operational reliability.

## Education

### University of British Columbia

Expected June 2024

*Master of Data Science*

*Vancouver, British Columbia*

### Universidad del Norte

March 2020

*Mechanical Engineering*

*Barranquilla, Colombia*

## Technical Skills

**Languages:** Python, SQL, R, MATLAB, Scala, Javascript, Java

**Technologies:** PyTorch, Tensorflow, Keras, scikit-learn, sktime, Spark, Hadoop, Git, PowerBI, Pandas, Polars, Matplotlib, Seaborn, Altair, Plotly, Docker, Azure, AWS, Apache Airflow

**Concepts:** Artificial Intelligence, Machine Learning, Neural Networks, API, Database Normalization, Agile Methodology, Cloud Computing, Data Engineering

## Projects

---

### **End to End Data Pipeline (ETL) using AWS (Data Engineering 2.0 project) | *Python, AWS, SQL, Redshift, PowerBI***

- Utilized the CMS.gov API to access and extract OpenPayments data. Orchestrated the seamless transfer and storage of the data within an Amazon S3 bucket, leveraging an EC2 instance for efficient execution.
- Leveraged an EC2 instance to execute the transformation process effectively, guaranteeing data integrity and compatibility for seamless integration and loading into the Redshift database.
- Following meticulous data extraction and transformation procedures, orchestrated the loading of the processed data into PowerBI from the Redshift database. This step ensures accessibility and enhanced usability for data analysts, enabling them to derive actionable insights and make informed decisions effectively.

## Certifications

---

- **Databricks Certified Machine Learning Associate**
- **DS4A: Data Science for All 6.0 (Graduated with Honours) & DS4A: Data Engineering 2.0**