

CARINA H. REBELLON

carinarebellon@outlook.com

+ (65) 98551334

Linkedin: <https://www.linkedin.com/in/rebelloncarina/>

Github: <https://github.com/rebellon-carina>

With over 15 years of experience in data warehouse design and analysis, focusing on requirements gathering, data modeling, design implementation and production support. A skilled systems analyst with hands-on experience and advanced knowledge of SQL and that specializes in ETL/ELT development, currently integrating big data engineering in the field of data science.

KEY SKILLS:

Tools and Software:

DOMO, Python, SQL, SQL Server Integration Services (SSIS), Informatica PowerCenter, Informatica Cloud, Oracle PL/SQL, Teradata, Snowflake, Unix Scripting, GitHub, SAP, SAP BI/BW, Cognos, Business Objects, Tableau

Big Data: Hadoop, HDFS, MapReduce, Pig, Hive, Spark, PySpark, Databricks, Docker

Microsoft Azure: Blob Storage, Data Lake, Databricks and Data Factory

AWS: EC2, S3, EMR, Kinesis, Hue, Athena, QuickSight

Machine Learning/Data Science: Supervised (*Regression, Classification*), Unsupervised (*Clustering*), Web Scraping, Natural Language Processing (*Sentiment Analysis*), Time Series (*ARIMA, SARIMAX*), Recommender Systems (*Content-Based Filtering, Collaborative Filtering, and Association Rules Learning*), Deep Learning (*TensorFlow*)

Data Engineering Projects:

End to end implementation of streaming live feed from Twitter using Kinesis Firehose, the tweets are stored in S3 bucket. Used Databricks to transform

the data, created pipeline for the sentiment analysis model and predict the sentiment of the data from Twitter. Visualized the predicted data using Athena and QuickSight.

Subscription and Churn Analysis (Proof of Concept) using Azure Databricks and Azure Data Factory.

Data Science Projects:

Exploratory Data Analysis on SAT and ACT examination from year 2019-2021. (*pandas, matplotlib, seaborn*) --[tableau](#)

Housing Price Prediction: Built regression models to predict the house price and give recommendation to prospective buyers based on property location and quality. (*feature engineering, LogisticRegression, Ridge, Lasso, ElasticNet*)

Natural Language Processing: Sentiment analysis and classification models to accurately label two different subreddit topics; web scraping using Reddit's PushShift API; stemming, lemmatizing, tokenizing the text from post to clean the data. (*CountVectorizer, TfidfVectorizer, DecisionTree, RandomForest, ADABOOST, GradientBoost, Voting, Bagging, NaiveBayes, SVM, StandardScaler*)

West Nile Virus (WNV) Prediction: Classification problem where we need to predict the presence of WNV based on location and weather features. (*DBSCAN, SMOTE, RidgeClassifier, ExtraTrees, RandomForest, GradientBoost, ADABOOST, TensorFlow, Pipeline, SelectKBest*) --[tableau](#)

Other projects in Databricks:

[Employee Retention](#), [Churn Analysis](#), [Customer Lifetime Value](#)

Employment:

Lead Senior BI Analyst/Data Engineer

Mar 2010 – Feb 2022

AVID (<https://www.avid.com/>)

Projects:

Involve in the design phase of the Enterprise Data Warehouse using Data Vault 2.0 model, integrations of cloud applications such as DBT, Snowflake, Fivetran and Velocity. **(Year 2022)**

The only data engineer resource in the Customer MDM (Master Data Management) project, designed and implemented data extraction, prepared tables and views and administered the roles and access in Snowflake. (*Tamr, Azure SQL, Velocity, Fivetran, Snowflake*) **(Year 2021)**

Technical resource for data integration for the AnytimeCollect project, a finance project to improve collection of accounts receivable. (*Velocity, Azure SQL, Azure Data Factory*) **(Year 2021)**

Built the foundation of the subscription and churn analysis dataset using RedShift SQL in DOMO, created dashboards for all the metrics and KPI related to license count, subscribers' growth and churn. **(Year 2019)**

Major project in 2019 was the retirement of the SQL/SSIS and Cognos where it was migrated to cloud technology (DOMO). Learned the tool quickly and contributed to the success of the project. **(Year 2019)**

Successfully retired the legacy reporting in AVID and built the ETL using SSIS from ground up, I have self-learned the ETL tool, created and designed the database with indexing for optimal performance in data loading and querying. **Year 2011-2012)**

Business Information Analyst

Nov 2007 – Mar 2010

3COM Asia Pacific RIM PTE LTD (acquired by HP)

Converted to permanent role and awarded with *Key Employee Retention Plan* (within 6 months of employment) after successfully implemented the Product Hierarchy project in the company.

Technology: Informatica PowerCenter, Oracle, Business Object

Senior Software Consultant Feb 2007 to November 2007

Apar Technologies Pte Ltd, Client: DBS

Technology: Informatica PowerCenter, Teradata

Certifications:

Big Data Bootcamp, SMU Academy (April 4-8, 2022)

MajorDomo Certification (2020)

Oracle PL/SQL Developer Certified Associate (2008),

Oracle Database SQL Certified Expert (2008)

Education:

Data Science Immersive, Singapore

General Assembly, Nov 2021 to Present
(Completion in June 2022)

Bachelor in Computer Data Processing Management, 1996-2000

Polytechnic University of the Philippines, Manila, Philippines

Achievement: **Cum Laude**

Status in Singapore: Permanent Resident