

# Carlos Marin

Houston, TX | 713-791-3494 | [carlosmarinjr1@gmail.com](mailto:carlosmarinjr1@gmail.com) | [Blog](#) | [GitHub](#) | [LinkedIn](#)

## Data Scientist | Data Analyst

Data Scientist with a background in clinical research, and a passion for using data and machine learning techniques to drive actionable business insights from startups to global businesses. Self-starter with experience in team building and data analysis, excelling at delivering complex information to business audiences, and delivering actionable results.

## DATA SCIENCE TECHNICAL SKILLS

- Programming: Python (Pandas, Numpy, Sci-Kit, statsmodels, Keras, TensorFlow, SQL, Pytorch), PySpark
- Data Visualization: Python (Seaborn, Matplotlib), Tableau, PowerBI, Microsoft Excel, Minitab
- A/B Testing, Hypothesis Testing, Multiple Linear Regression, Machine Learning, NLP, Deep Learning
- Software Engineering: JavaScript, React, Express.js, Django, HTML, CSS

## TECHNICAL PROJECTS

### Inflation Impact Analysis using Python | [Github](#)

Developed multiple machine learning models, evaluated performance, and predicted customer turnover and causes.

- Implemented odds ratio algorithm to evaluate best constraints on predictors. Programmed GridSearchCV to find best parameters for each model.
- Utilized multiple machine learning techniques including Logistic Regression, KNN, Decision Trees, and Random Forests.
- Interpreted each model performance using Classification Report and Confusion Matrix.

### Spotify Music Recommendation using Spotify API | [Github](#)

Using Spotify API, accessed Spotify playlist to create Music Recommendation System

- Programmed Spotify API to make POST request to make authorized request to retrieve music data.
- Developed function using Spotipy, to collect music data from Spotify on selected Spotify playlist.
- Implemented a music recommendation system based on Audio Features using cosine similarity.

### Disaster Text Classification using BERT | [Github](#)

Using Natural Language Processing (NLP), conducted analysis of positive and negative texts classification.

- Programmed Transformer BERT model for Text Classification
- Incorporated Logistic Regression and GridSearchCV with Count Vectorizer and TFIDF Vectorizer to compare machine learning model performance.
- Developed WordClouds as a data visualization technique to visualize most frequent words of Targets.

### Diabetic Retinopathy Image Classification Using CNN | [Github](#)

Using Tensorflow.Keras Sequential Deep Learning Model, specifically CNN, for image classification.

- Programmed CNN deep learning model to distinguish between different classes of Diabetic Retinopathy.
- Interpreted model evaluation using Classification Report and Confusion Matrix.
- Implemented MobileNet in a transfer learning model as a fixed feature extractor and fine tuning.

## PROFESSIONAL EXPERIENCE

Clinical Research Coordinator - Heart Failure, **Houston Methodist Hospital**, Houston, Texas 09/2021 - Present

- Conducted the coordination of numerous heart failure studies including both drug, device, and registry studies.
- Screened subjects from the daily clinic schedule and confirmed eligibility of potential study subjects.
- Prepared and processed all documentation through Institutional Review Board (IRB), including submissions, continuing reviews, amendments, and adverse event reporting.

## EDUCATION

**Flatiron School, New York, NY**

Immersive Data Science Bootcamp program

January 2023

**University of Houston - Downtown**, Houston, TX

Bachelor of Science, Biological and Physical Sciences

May 2017