**LUCAS REYES**

**DATA SCIENTIST**

**E-mail** **:** reyeslucas@gmail.com **Phone**  **:** +1 571 330 30 30

**Linkedin :** [linkedin.com/in/lucas\_reyes](https://www.linkedin.com/in/lucas_reyes) **Address** **:** USA

**GitHub :** github.com/Lucas\_reyes

**Tableau Public :** public.tableau.com/app/profile/reyeslucas

**SUMMARY**

Cleaning up the live scraped data from the site with the Auto scout End to end real life regression project, a car sales platform, EDA process (check data quality and feature engineering to create new features, clean, organize, analyze, summarize, and visualize data) After the implementation of the EDA process, shortening the sales period by suggesting a price with predictive analysis (Python Scikit-Learn) while selling the vehicle to the new person who will come to the site. And deploy on AWS using Flask or fluent.

Experienced in IT and Machine Learning, Artificial Intelligence, Statistical Modeling, Data Analysis, Predictive Analysis, Data Manipulation, Data Visualization, and Business Intelligence. Proficient at building robust Machine Learning & Deep Learning models.with Scikit-Learn, Keras and Tensorflow. Python data extraction and data manipulation, and widely used python libraries like NumPy, Pandas, Seaborn, and Matplotlib for data analysis.

**WORK EXPERIENCE**

**Nioya Tech LLC, Virginia(Remote)** Mar 2021 – Present

**Data Scientist**

* Skilled in developing models using techniques such as Logistic Regression, Linear Regression, Random Forests, Gradient Boosting, Cluster Analysis, Association Rule, Bayesian Algorithms, Neural Networks, Deep learning frameworks such as TensorFlow, Keras.
* Worked on data projects, improving classification results, creating data pipelines and exploratory visual presentations to clients.
* Applied methodologies, Python; NumPy, SciPy, Matplotlib, Pandas, Seaborn.
* Hands-on experience in HTML, GIT, GitHub, Jira, SQL, Tableau, and Linux.
* Established advanced customer analytics using customer segmentation, k-means, cohort analysis, and RFM.
* Extracted knowledge and insights from different data types using scientific methods, algorithms, and systems and got the essential skills expected from a data scientist.
* Built data pipeline for features extraction, data cleaning, and features scaling using Pandas and NumPy.
* Gained the leadership and communication skills necessary to deliver measurable and tangible results.
* Extracting and transforming data, Model deploying flask, streamlit, AWS
* Worked on many projects such as Customer segmentation, Churn Prediction, Sentiment Analysis, Image Classification, Sentence similarity, Document classification, Car&House price Prediction, Keras Regression&Classification, Customer Retention with SQL and Tableau, RFM, and Cohort analysis
* Hands-on experience with accessing, cleaning, and converting secure clinical datasets for preprocessing to build Natural Language Processing models.
* Experienced with applying pre-trained embeddings and the state-of-the-art (SOTA) models to the custom datasets and tuning their parameters as the best param models. (Involved in building clinical NER models like the de-identification model.)
* Developed Sequence Classification, Language Modeling, Text Generation, Named Entity Recognition, Summarization, Translation NLP models, and pipelines using the SOTA architectures.
* Built OCR pipeline for converting images to text by using Open-CV, Tesseract, and Easyocr. (The last project was detecting handwritten text from the image and pdf documents.)
* Developed multi-step LSTM time series forecasting models for power usage, by using long short-term memory recurrent neural networks for multi-step time series forecasting of household power consumption.
* Applying real-world data cases for identifying, analyzing, and interpreting trends and creating interactive dashboards using Tableau to understand insights of the business.
* Extracting, processing, and analyzing complex datasets using visualization, and analytics tools.
* Developing strong programming skills in Python as well as an understanding of open-source frameworks.
* Deriving predictive models using appropriate statistical techniques.
* Ensuring the data that supports the project recommendations is verified, analyzed, and modeled effectively in collaboration with other team members using Enterprise Decision Analytics (EDA) software.
* Collecting and cleaning data with MS SQL Server using advanced queries, and processing large datasets with Python.
* Hands-on experience in statistical research techniques such as Linear/non-linear regression, sample size estimation, and statistical testing.
* Gained the leadership and communication skills necessary to deliver measurable and tangible results to stakeholders.
* Great communication skills and the ability to present technical topics to a non-technical audience.
* Well-organized, self-motivated team player; aptitude for project management.

**XXX Company Sep 2009 – Mar 2021**

**Electronic Engineer**

* Created and proofread presentations for top-level executives,
* Scheduled and headed technical support team meetings and took care of technical support administrative functions,
* Successfully supervised the overall activity and interpersonal relationship among members of technical branches,
* Worked in large projects interacting with several teams within the project. Prepared close-up analysis reports regarding problem-solving and risk management techniques implemented during the project.

**PROJECTS**

* Customer Review Sentiment Analysis (NLP): Retrieved shopper comments on consumer products from shopping websites and utilized machine learning models to extract keywords from comments and categorized them by sentiment and features they describe.

Skills: Tokenization, Noise Removal, Lexicon Normalization, *Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest,* and *Ada Boosting.*

* Customer Segmentation: Segmentation of customers using unsupervised data and make some suggestions to clients about segmented customers.

Skills: K-means Clustering, Dendrogram, Agglomerative Clustering, 3D Visualizations,

* AutoScout Car Price Prediction (Using Real-Life Data) : Building price prediction model and deploy using features regarding different make & models.

Skills: Feature Engineering and Analysis, Pandas, Seaborn, scikit-Learn, on AWS Flask or streamlit.

* Tree Types Classification: Predicting types of trees that grow in national forest district using the features shadow coverage, distance to nearby landmarks, soil type, etc.

Skills: Support Vector Machine(SVM), XGBoost, Random Forest, Decision Tree, KNN, Yellowbrick, Seaborn

* RFM Customer Segmentation & Cohort Analysis Project: Consists of 4 different projects that contain different scenarios. RFM Analysis, K-Means Clustering, and Customer Segmentation are applied for customer segmentation analysis.

Skills: Pandas, seaborn, Matplotlib, K-Means, Data Cleaning, Data Visualization, and Exploratory Data Analysis capabilities.

* Employee Churn Analysis Project: This project classification model for a variety of business settings. What is Employee Churn?, How it is different from customer churn. Implement classification techniques. Make predictions with supervised algorithms. At the end of the project, deploy the model using Streamlit.

Skills: Exploratory data analysis and visualization using matplotlib and seaborn, model building and evaluation using python scikit-learn package. Support Vector Machine(SVM), XGBoost, Random Forest, Decision Tree, KNN, Yellowbrick, Seaborn

* Fraud Detection Project: Dataset presents transactions made by credit cards in September 2013 by European cardholders occurred in two days, where it has 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. Explore the data to get pattern of the fraud and build a supervised model to identify them. Lastly, deploy the best model by using Flask.

Skills: Logistic Regression, Random Forest, Neural Network algorithms and SMOTE technique. Seaborn, Matplotlib , and Yellowbrick.For deploy  Flask API.

* Demand Prediction Project: It is aimed to reduce traffic congestion, noise, and air pollution. In this project, the goal is to predict the number of future bike shares given the historical data of London bike shares. So this case should be handled as a time series problem with Deep Learning.

Skills: Deep Learning(Bidirectional LSTM, RNN), Pandas, Seaborn, Matplotlib , and Yellowbrick.

* Image Classification Project: Build an image classifier with Convolutional Neural Networks for the Fashion MNIST dataset. This data set includes 10 labels of different clothing types with 28 by 28 grayscale images. There is a training set of 60,000 images and 10,000 test images.

Skills: NumPy, Pandas, Seaborn, Matplotlib, Deep learning(CNN)

* Analyze US Citizens (EDA Project): Analyzing the characteristics of individuals according to income groups. Data Cleaning, Feature Engineering, and Data Analysis.

Skills: NumPy, Pandas, Matplotlib, Seaborn

* Car Price Analysis (Using Real-Life Data): Analyzing data with Exploratory Data Analysis (EDA) process to get useful insights about car prices. Preparing the dataset for an ML model to use it to predict car prices appropriately. Data Cleaning, Feature Engineering, and Data Analysis

Skills: Numpy, Pandas, Matplotlib, Seaborn.

* Bike Demand Visualization Project: Analyzing and visualization of bike demand using the historical data of London bike shares. Data Cleaning, Feature Engineering, and Data Analysis

Skills: Numpy, Pandas, Matplotlib, Seaborn.

* Tableau Covid-19 Tracking Dashboard: A dashboard was prepared to monitor the daily number of Covid-19 cases and deaths worldwide. In order to obtain daily data, a server connection was established via the web data connector. After the data preparation and data manipulation processes, the dashboard was prepared and this dashboard, which was prepared via embed code, was made to be followed live on the website (google sites).

Skills: Data Preparation, Data Manipulation, Data Visualization, Tableau.

* E-Commerce Data Analysis Project: Used E-Commerce data; In addition to customer segmentation, churn analysis, and customer retention rate applications, which have wide applications in the industry, different data analysis applications are exemplified.

Skills: Data Preparation, Data Analysis, Data Manipulation.

* Data Analysis with SQL: Analysis of the sales and delivery system of an e-commerce company. Skills: SQL
* Database Design of a University: An example university registration data model and database. Skills: ER Diagrams, SQL
* Police Kill Activity: USA fatal police shoot analysis using the visualization tools. Skills: Pandas built-in, Matplotlib, Seaborn

**TECHNICAL SKILLS**

Data Analysis | Business Intelligence | Machine Learning | Natural Language Process (NLP) | Deep Learning | Statistics | Numpy | Pandas | Matplotlib | Seaborn | Yellowbrick | Sci-kit Learn | Tensorflow | Keras | NLTK | Python | SQL | HTML & CSS | MS SQL Server | Tableau | Google Sheets | Linux | Git | Github | Jira | Agile Methodology

**SOFT SKILLS**

Confidence | Leadership | Time Management | Problem-Solving Abilities

**EDUCATION**

**Yale University - Electronic Engineer** Sep 2005 - Jul 2009

**CERTIFICATIONS**

* Fundamentals of Visualization with Tableau, Coursera Nov 2021
* Applied Machine Learning in Python by University of Michigan, Coursera Oct 2021
* Applied Text Mining in Python by University of Michigan, Coursera Sep 2021
* Natural Language Processing With Classification And Vector Spaces, Coursera Aug 2021
* The Complete Google Sheets Course: Beginner to Advanced, Udemy Jul 2021

**LANGUAGE COMPETENCIES**

**Turkish**: Native

**English**: Advanced