# Tran Quoc Bao

## Data Analyst Intern

### CONTACT

📞 +84 777 257 131

🌐 https://bao-portfolio.vercel.app

✉️ baotran310702@gmail.com

in baotran310702

⌗ baotran310702

### EDUCATION

**University of Information and Technology - VNU**

- Bachelor of Information System
2020-2024; GPA: 7.5/10

### COURSES

**Google Data Analyst Certificate Coursera.**

**Advanced SQL Certificate by Kaggle**

### SKILLS

- MS Excel (VBA in writing Macros and VLOOKUP at basic level)
- MS Power BI (Visualization & Dashboard)
- Python (numpy, pandas, matplotlib, pytorch, selenium ...)
- Query Language (SQL, MySQL, Oracle...)
- English (Immediate - Upper)

## ABOUT ME

I am a junior student at the Vietnam National University of Information and Technology. I have a strong passion for understanding how data drives insights and decision-making processes. With a keen interest in the field of data analysis, I am eager to expand my knowledge and skills in this domain. I hope to be a part of your team.

## PROJECTS

### Analyzing the behavior of tenants in New York 2019 using KMean - a clustering algorithm.

- Data preprocessing with Python (drop empty, delete duplicate...), then drop all columns that have missing values more than 10%. If the columns have too much missing values, it's useless to analyze. Columns that contain id, name is useless also be dropped.
- Using heatmap to find out the correlation attributes. The result shown that all attributes left have weak correlation.
- Basic statistic with them to find out mean, median, quartile and standard deviation.
- Using one-hot-encoding with qualitative columns. Then using PAC (Principle Component Analysis) to reduce dimensions.
- Calculate SSE (Sum of Squared Error) and Dendrogram to visualize the cluster. The result show that there are 3 clusters in this dataset base on reduced using PAC.
- Visualization their best correlation between qualitative and 3 clusters with box-plot chart and show the neighbourhood_groups in bar chart in clusters.

*Link Github: https://github.com/baotran310702/Crypto_Price_Forecasting*

### Building a data warehouse for dataset AB_NYC_2019 (Airbnb at New York City in 2019) using Visual Studio and SQL Server (Team work)

- Design Star Schema model for dataset.
- Build ETL pipeline to database using SSIS in Visual Studio.
- Prepare Server Integration Services and build data warehouse.
- Run and deploy data and its relationship to SQL Server Database..
- Using Analyst Services to make reports and data query by MDX.

*Link Github: https://github.com/baotran310702/SSIS_OLAP*

### Using ARIMA and Hybrid ARIMA & Random Walk Models (ARIMAR) to predict price of cryptocurrency (ADAUSDT) in Binance Exchange.

- Preprocessing dataset, drop useless columns like volume, ignore... and keep usable columns like OLHC and timeseries, find out the correlation attributes.
- Dataset is fetched by using API provided by Binance and save to csv file.
- Using ADF-test to check if timeseries is stationarity or not. Then applying differencing method, PACF and ACF method to find the best parameters for ARIMA model.
- Train model with different ratios by split dataset to 7-2-1 & 6-4-4 (train - validate - test) and evaluate them with RMSE (Root Mean Squared Error) value to find best train data for predicting.
- Predict price with ARIMA model. With ARIMAR, using random function to random noise and use it to predict next step of data then combine the predict result of ARIMA and Random Walk using smoothing moving average.
- Visualizing the predicted values and train - validate -test data with matplotlib.

*Link Github: https://github.com/baotran310702/Crypto_Price_Forecasting*

## ACTIVITIES

- Participated in Hackathon UIT Web - Game Competition.
- Member of Content Writer Team in Software Engineering Study Group.