

DONG MING ZHEN

ADVANCED ANALYTICS AND DATA SCIENCE

SUMMARY

Dong is a data scientist who practiced marketing for over two years with the intent to cultivate the relationship between businesses and their target markets. He designed brand guidelines, call to action emails, and website copy to communicate business values to consumers. During this process, he learned about data science and practiced a combination of programming, engineering, and analysis skills that can help businesses find data driven insights to make better consumer decisions and track the results over time.

EXPERIENCE

Metis · Remote

Data Scientist · Mar. 2021 to July 2021

Completed an immersive 10 week accredited data science bootcamp to perform EDA and trained machine learning models on real-world datasets, then communicated results using presentation best practices. Projects include:

Where to fish in New York State?

- Created a data pipeline that scraped, cleaned, and merged 30,000 forum posts from New York's top two fishing forums to aid the New York Department of Environmental Conservation address the top ranked request from fishermen/anglers
- Used spaCy and RegEx to clean and preprocess the text data through string extraction, tokenization, lemmatizing, and stop words
- Created a document term matrix with a TDIF vectorizer, then used NMF to find relevant fishing topics from reoccurring terms
- Extracted strongly defined topics to recommend fishing location and gear by season

Mitigating Delta Airlines' Operational Losses

- Constructed a data pipeline to extract the names of U.S. airports and then used Selenium to automatically extract airline data from the Bureau of Transportation Statistics database.
- A larger csv file with 200 million rows was taken from IBM and chunked to find the relevant rows, then merged with the first dataset
- Selected only predictive features that were known ahead of time before a flight departs, and invented new features to help with the prediction such as scheduled number of flights a plane has in the day
- Created a stratified cross validation pipeline to address class imbalance and trained the data on logistic regression, random forest, and xgBoost models
- The models' low recall scores revealed that there are missing predictive features that need to be collected for the next iteration.

Increasing Readership for the New York Times

- Extracted New York Census data by calling it's API, scraped New York zip code and location data using requests and BeautifulSoup
- Propose training a logistic regression model on existing subscriber demographic data, then testing it on Census demographic information obtained by calling its API, to identify which New Yorkers are likely to be a subscriber or not
- Created a dashboard to capture income, education, and occupation trends by age groups and gender. Mapped the density of age groups using location data to help New York Times identify different ways to segment and target the New York demographic
- Addressed potential data ethic issues with the model such as design bias and offered solutions like giving more weight to features that describe readers that New York Times is interested in

Predicting a Player's Score

- Scraped NBA team statistics, roster, and player box score data from the 2017 - 2019 seasons using basketball-reference.com with BeautifulSoup
- Identified the features that are important to predicting a basketball player's Fanduel score including offensive and defensive tendencies; created features from existing data to proxy for missing attribute features
- Set up a cross validation train test split to score a linear regression model across five fold
- Normalized skewed features using log transformation, and adjusted the features to a polynomial of degree 2 to train a linear regression model that made predictions with a mean absolute error of only 4.9

Open Your Next Business Here

- Cleaned messy MTA turnstile and location data by removing null values and fixing inconsistencies to help identify a profitable location for a local snack business
- Grouped stations outside of Manhattan with high entries in the morning and exits in the afternoon to create a commuter station dataframe
- Used Matplotlib and Seaborn to visualize the top commuter stations and the ones with consistent traffic over time
- Recommended Junction Boulevard in Queens if the company is interested in the busiest commuter station or 77 St in Brooklyn if they want a steady stream of commuters during morning and evening meal hours

Amorium LLC · New York, New York

Marketer · 2017 to 2018

B2B jewelry company that specializes in delicate handcrafted sterling silver pieces with trendy designs. Tasked with helping to develop a marketing strategy to retain retailers while finding ways to direct the target market to the Amorium website.

- Developed branding guideline to maintain consistency in brand communications
- Created search optimization guideline and campaign to improve the search results of Amorium and its many product pages; wrote product descriptions based off relevant high cost per click and low competition keywords and its latent semantics. Wrote blog posts and identified back linking opportunities
- Managed email marketing campaigns on Salesforce and Mailchimp to attract and retain customers
- Assisted management with daily priority tasks such as designing PowerPoint presentations, creating product catalogs, and customer assistance

CONTACT

✉ dong.zhen.marketing@gmail.com

☎ 917.838.8737

📍 Brooklyn, NY 11228

in /in/dong-ming-zhen/

👤 Dong-Zhen

SKILLS

PROGRAMMING LANGUAGES

Python

SQL

MACHINE LEARNING

Regression

Classification

Clustering

Dimensionality Reduction

Topic Modeling

TOOLS

Pandas

Numpy

BeautifulSoup

Selenium

Scikit-learn

Natural Language Processing

Excel

VISUALIZATION

Tableau

Matplotlib

Seaborn

Plotly

EDUCATION

Baruch College

Bachelors of Business

Administration Marketing

ACTIVITIES

Bay Ridge Toastmaster Club ·

Presentation Pathway

Apr. 2021 to Current

Practice public speaking skills such as story telling with a supportive and experienced speaking group