# CANDICE YUN CHEN

**Answering Business Questions using Data Science**

📞 949-232-3291          @ cyycan@gmail.com          📍 Marina Del Rey, CA

## EXPERIENCE

### Sr. Manager Risk and Compliance  Data Science

**Protiviti Inc.**

📅 08/2021 - Present          📍 Location

Protiviti Inc. is a global consulting firm headquartered in Menlo Park, California that provides consulting in internal audit, risk and compliance, technology, business processes, data analytics and finance.

- Ongoing mentorship for senior and manager level consultants for the department
- Actively participate in research for patent application for the AI for health care.
- **End Client: AstraZeneca**
- Led/Performed living patient journey analysis by built rule based model using claim dataset via azure databrick
  **End Client: JP Morgan Chase**
- Validate anti-money laundering machine learning model using RStudio
- Support ongoing audit effort via azure databrick
- Use lime and Shap to interpret text classification machine learning model via azure databrick
  **End Client: TR Price**
- Travel Expense Analysis and outlier detection using jupyter notebook
  **End Client: Affirm Inc.**
- Model validation for XGBoost Model via azure databrick
  **End Client: Kimberly Clarke**
- Assist data audit on third party vendor and investigate in fraud and forensic accounting
- Which of your achievements match the job you're applying to?

### Lead Data Scientist

**Cognizant Softvision**

📅 07/2019 - 07/2021          📍 Marina Del Rey, CA

Consulting company

- **End Client: Kaiser Permanente**
- **Team: Data Engineer**
- **Title:** Lead Data Scientist
- Performed large-scale data analysis and develop effective statistical models for NLP, segmentation, classification, optimization, time series, in python etc
- Analyzed and Automated multiple dashboards using SQL, Python, and Tableau for varieties of stakeholders such as UIUX, Sales, Product Managers, leadership
- Investigated data characteristics, complexity, quality, and meaning through visualizations
- Defined performance metrics; identify trends, outliers and drivers of outcomes in python pandas
- Developed mitigation strategies for potential problem areas
- Created and presented results to executives and other stakeholders with finding in clustering search phrases and top viewed documents to find out what people intend to search and segment common behaviors people do analysis or search engine using NLP techniques in python

## KEY ATTRIBUTES

### Business Needs Driven

Keeping a constant pulse of the business environment, needs and shifting priorities to create the most relevant deliverable.

### Analytical

A knack for truly wanting to understand how things work, finding root causes, correlations and deriving a deep satisfaction proving out a hypothesis.

### Curious

Always eager to learn about ideas, solutions and topics whether they are directly applicable to my field or beyond.

### Passionate

Capable of turning dry analysis into an exciting story that influences the direction of the business with diverse teams

## ACHIEVEMENTS

### Implemented Tableau-based Enterprise SaaS Solution

Lead the conception and implementation of a high-availability real-time analytics solution based on Tableau Server, which is fully integrated with a SaaS platform (SSO) servicing 300+ users

### Data Management - 70 projects - 100% On-Time Completion

Managed hundreds of data exchanges with health plans over the course of 3 yrs

### Increased Search and Play Rate by 10% across 20 Million customers

Optimization strategy drove customer engagement and usage, both key ingredients for customer retention

# EXPERIENCE

## Principal Data Scientist
**Undisclosed (Stealth Mode)**

📅 09/2018 - 06/2019  📍 Malibu, CA

- Outlier detection using dbscan, median absolute using Splunk
- Run forecasting with LSTM models using Keras library in python
- Work cross functionally to define problem statements, collect data, build analytical models and make recommendations
- Evaluate business analytics strategies and determine analysis goals
- Design, develop BI solutions for Health Care Client using, Amazon S3 & Redshift

## Data Lead
**Bank Of America**

📅 09/2017 - 09/2018  📍 Thousand Oaks, CA

Bank - Credit Loss Forecasting Analytics

- Train and coach a team of 15 members machine learning e.g. clustering, dimension reduction from step forward selection
- Train and coach a team of 15 to use BI tools to replace excel as part of the BI as Service Strategy
- Led, build, tested and automated ETL for 5 projects in SQL, Python/Pyspark/Hadoop and delivered on-time
- Created mortgage risk model prototype using deep learning tensorflow in python with panda dataframe
- Cluster price tiers to further assist business decisions using KMean in python with panda dataframe connect with MS SQL db using SQL lite python library
- Automate and build data pipeline in MS SQL for distress discount and cost to sale dashboard in Tableau
- Run impact sensitivity analysis for model validations in python pandas dataframe and SQL lite and visualize results in tableau

## Data Scientist
**AT&T Big Data**

📅 05/2016 - 06/2017  📍 El Segundo, CA

AT&T Entertainment (formerly DirecTV) is a provider of digital entertainment services in the United States and the world. BI Department

- Increased Search and Play Rate by 10% across 20 Million customers by
  - created, automated and monitored actionable and meaningful KPI
  - created, analyzed and presented behavior models
  - created and presented actionable insights to increase customer engagement, retention by regression and hypothesis testing to varieties stakeholders (UIUX, product team, engineer etc.)
- Created predictive models for customer engagement, retention and segmentation in python via RCloud and visualize results in tableau
- Design, develop BI solutions using, Amazon S3 & Redshift.
- Profiled raw data sets across platforms and develop KPI/dashboard to measure product performance

# MY WORKDAY
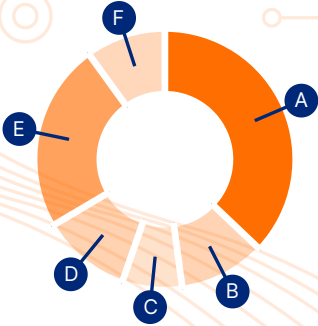


- Ⓐ Stakeholder Interactions
- Ⓑ Analytics Design / Prototyping /Data Visualization
- Ⓒ Modeling
- Ⓓ Data Discovery and Cleansing
- Ⓔ Reading / Planning / Learning
- Ⓕ Team Presentation

# SKILLS

**Data Analysis/Analytic Problem-Solving**

**Data Visualization/Story Telling/Communication/PM**

**Visualization Tool (Tableau/MicroStrategy)**

**SQL**

**Big Data Platform (Hadoop, Spark)**

**Python/R**

# CERTIFICATES

**Deep Learning**
Coursera

**Data Science John Hopkin**
Coursera

**AI Business Strategy and Application U.C. Berkeley**
In-progress

# EXPERIENCE

## Senior Data Scientist and BI Lead (Full time)

**HEALTH DATA VISION**

📅 10/2010 - 03/2016    📍 Burbank, CA

Health Data Vision, Inc. provides a leading SaaS platform for health plans and service providers to perform mission-critical medical record review processes (MRR) in the areas of quality and risk adjustment

- Collaborate with the executive management team to rapidly prototype business dashboards and KPIs
- Create real-time dashboards to track all operational metrics against goals and service level benchmarks
- Discover and analyze transactional and metadata sets to support optimizing operating models with dashboards for billing, cost attribution, SLAs, etc
- Perform quantitative analysis to forecast project and sales trends
- Identify gaps in patient care in the claim, diagnosis, and Rx data set
- Find provider fraud using claim, member and prescription data for clients such as United, Aetna etc

# EDUCATION

## B.S. Information Computer Science

**University Of California, Irvine**

# REFERENCES

**References available upon request**

Reference Contact