
Pacific Life + DataMine
Because Tomorrow Matters

About Pacific Life

For more than 150 years, Pacific Life has helped millions of individuals and families with their financial needs through a wide range of life insurance products, annuities, and mutual funds, and offers a variety of investment products and services to individuals, businesses, and pension plans. Pacific Life counts more than half of the 100 largest U.S. companies as its clients and has been named one of the 2022 World's Most Ethical Companies® by the Ethisphere Institute.

Our Mission

To provide financial security through products and services that stand the test of time.

Our Vision

To be the company of choice providing financial security and well-being through industry-leading innovation.

Our Core Values

People; Accountability; Customer Focus; Integrity; Financial Strength; Innovation; Community.

The Product Management Team

The Product Management team provides innovative solutions to the design of life insurance products that makes life insurance easier for those who need financial protections. We build cutting edge actuarial and data analytics technology that provides our company with unique competitive advantage. We invest in bright, agile, and diverse talent to contribute to our mission of innovating our business and creating superior products.

Years of history

150+

Providing social good is who we are

Company Asset

$209B

Strong financials is how we thrive

Policyholders

1M

Protecting our policyholders is what we do

What you will get

- Hands-on machine learning application on real-world business data
- Develop business intuition from analyzing millions of data records
- Experience in working with commercial tools including Azure DevOps, Snowflake, MS SQL Server, etc.

Who we are looking for

We are looking for students who are curious and proactive in finding answers in the data and not daunted by working with large messy datasets.

Students who are familiar with Python and the concept of unsupervised learning will find this project more suitable to their skillsets. Basic understanding of SQL syntax will also help.

Project: Identify Policyholder Needs with Unsupervised Learning

Project Description:
Policyholders have all kinds of financial protection needs, including loss of income, retirement, and leaving a legacy. These needs can vary depending on the stages of life, socio-demographics, and many other factors. The team will be asked to identify the niche of these needs by using their statistical knowledge and Python skills.

The students will work with large sets of proprietary data from different sources in all kinds of formats. They will be expected to build efficient data pipelines to consolidate, clean, and manipulate the data for analysis. Students will develop business intuition along the way, where they will be asked to think along with us to identify new market opportunities.

Keywords: Machine Learning, Python, Life Insurance

Tools/Skills that will be used/learned:
- Primary: Unsupervised Learning techniques, Python, SQL, MS SQL Server, Snowflake, Git Repos
- Secondary: Insurance business concepts including the design of a life insurance policy, the concept of premiums, death benefits, cash surrender values, and how they protect the policyholders.

Preferences:
- An interest in learning about the life insurance industry
- Proactive in asking questions and in conducting independent research to close knowledge gaps
- A burning curiosity to ask why on the data and drill deep to find the answer

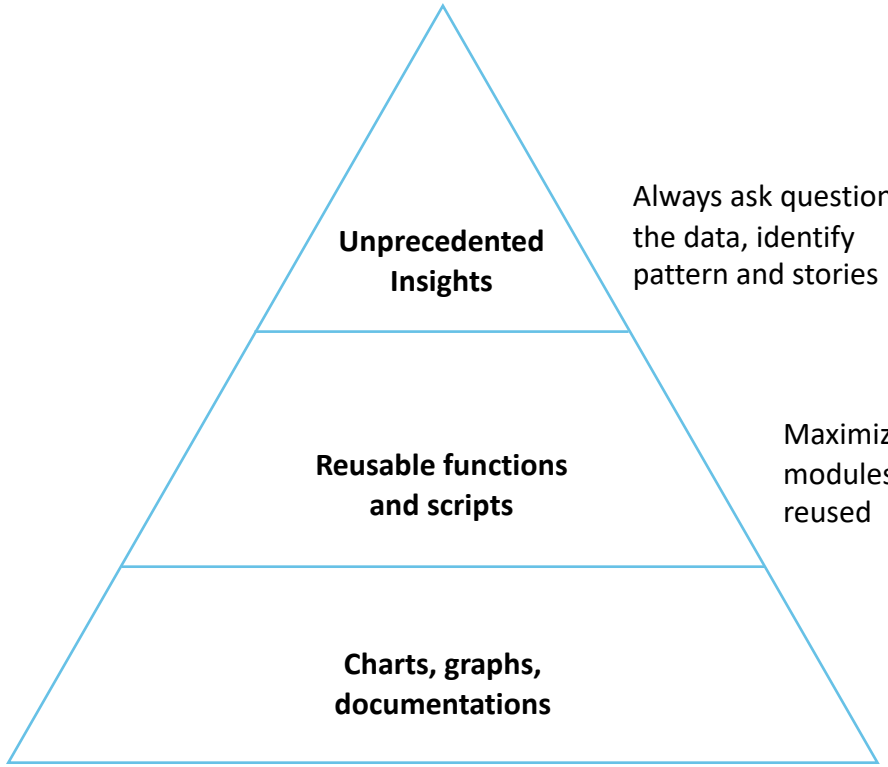
Citizenship: Open to all students

Expectations

- Be proactive and independent in your search for solutions
- Prove your value by educating your corporate partner on what they don't know
- Always come back with more –more analysis, more ideas, more information
- Communication – consolidate your larger questions send to us on Thursday evening.
- Use the Team chat for quick questions and sharing ideas.

Use the Value Pyramid to Always Deliver

Exploration of data is like drilling for oil. There is risk that nothing will come out of it. Therefore, keep in mind the value pyramid to always deliver something useful.



Always ask questions on the data, identify pattern and stories

Maximize the use of Python modules and functions that can be reused

Document every findings, even if not significant, at least they serve as a proof for path attempted.

Break-down of Fall Semester Work Plan

Sprint 1: Preparation and Setups

- Set up Amazon Workspace
  - install your favorite python editor (VSCode, Anaconda, etc.)
  - Link to Azure DevOps for Git repository
- Understand what an illustration document is by looking at the sample illustration pdf.
- Work with a sample json file to extract fund allocation information

Sprint 2: Exploration Data Analysis

- Extract from the json files first order variables and graph their distribution by counts, face amount, and premium
  - For example: first order variables are: by distribution channel, Age groups (0-34, 35-65, >65), NLG, LTC, Listbill, Withdraw indicator, major products,
  - Examine the distribution of first-order variables distribution across producers
- Within the first-order variables, graph the distribution of 2nd order features, which is everything else.
- Document all combination of charts

Sprint 3: Explore the use of unsupervised learning on dataset

- Separate into two workstreams, one conducts clustering analysis on policyholders, whereas another applies to producers.
- With the intuition developed from sprint 2, begin apply unsupervised learning on dataset, including within each of the 1st order variables.
- Examine different unsupervised learning techniques, their scoring mechanism, and their appropriateness on the different subset of the data

Break-down of Fall Semester Work Plan

Sprint 4~6: Reiteration of Unsupervised learning on the dataset

- Based on the feedback of each iteration, improve the process for the subsequent sprint
- Examine the cluster involvement over time
- Examine the cluster distribution for inforced policies and submitted policies
- Introduce new data source – i.e. health data as they come available
- Ad-hoc EDA analysis and introduce new features

Sprint 7: Wrap-up and Documentation

- Fix bugs and standardize all scripts for reproducibility.
- Document the use of scripts and findings
- Prepare for year-end presentation