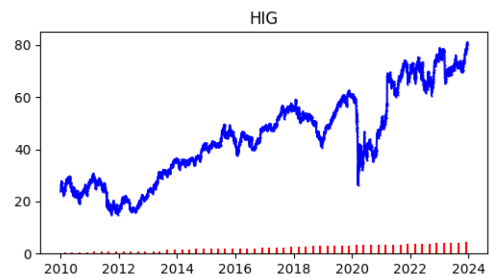
Part A - Finance

The Hartford Financial Services Group, Inc. is a Fortune 500 company offering property & casualty insurance, group benefits and mutual funds. The company was founded in 1810. In the good old days, the company provided home insurance for Abraham Lincoln and provided bond for the construction of the Golden Gate Bridge.

The file `HIG_price.csv` contains daily closing prices of the common stock (whose symbol is HIG). The file `HIG_dividends.csv` contains the ex-dividend date and the amount of past dividends.

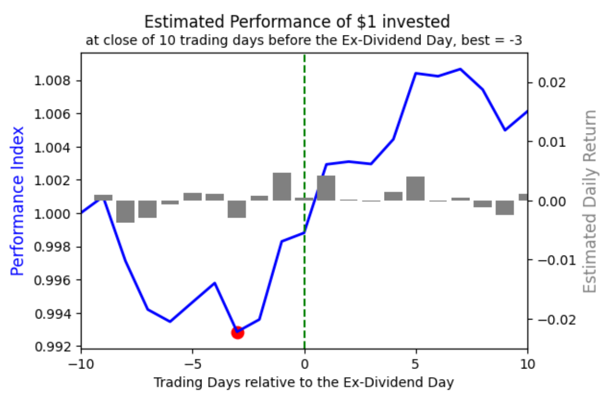
[A] Code in python to produce a chart of the price and dividend history of HIG, similar to the one below. The blue line is the stock price. The red bars indicate the ex-dividend dates and are proportional to the amount of dividends to offer a sense of dividend growth over time.



The Hartford paid $1.745 per share in dividends to its common stock holders over the past year. I am interested in receiving its dividends in 2024.

To reduce market exposure, I plan to buy the stock within 10 trading days of each ex-dividend date at the close and sell the stock on or shortly after the ex-dividend date.

Based on the historical performance of HIG, it appears that the best timing to buy the stock is 3 trading days before the ex-dividend date.



In the above chart,

- Day t is relative to the ex-dividend date. Day $t = 0$ is the ex-dividend date. Day $t = -1$ is one trading day before the ex-dividend date. Day $t = +1$ is one trading day after the ex-dividend date.

- The vertical bar at t is the average return for Day t, averaged over all dividend instances with sufficient price history for this analysis.

$$R_t = \frac{P_t}{P_{t-1}} - 1, \text{ for } t \in [-9, 10] \text{ and } t \neq 0$$

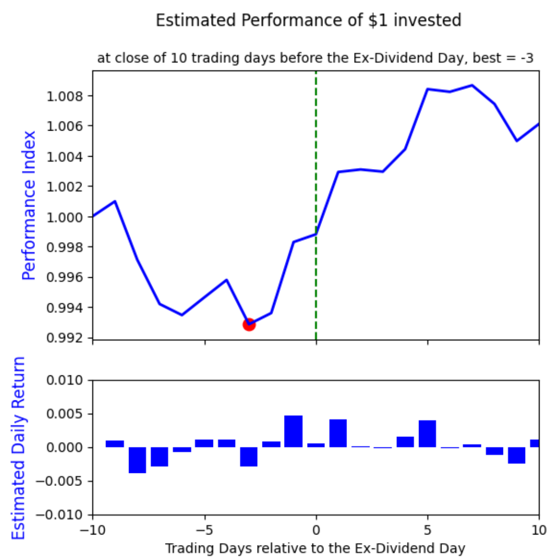
$$R_0 = \frac{P_0 + D}{P_{-1}} - 1$$

where P_t is the closing price for Day t and D is dividend per share for that dividend instance.

- The solid line is calculated as

$$V(t) = V_0 \prod_{t=-9}^{10} (1 + R_t), \text{ with } V_0 = 1$$

[B] Code in python to validate the optimal purchase timing.
[a] Let your code print out the value of R_t for $t \in [-9, 10]$
[b] Let your code reproduce the chart on page 1.
[c] If it is challenging for you to complete [b], please let your code reproduce the single chart below.



Part B - Insurance

The file `vehicle_claims.csv` contains vehicle accident claims and select characteristics of drivers and vehicles for an insurance portfolio in a specific year, where

- Sex: M = male, F = Female, U = data not collected
- VehicleType: A = automobile, M = motorcycle, T = truck, etc
- Claims_Count: total number of claims for that vehicle
- Term: the fraction of the year that the insurance policy is in effect
- Discount: discount based on the past claims of the policyholder
- AgeCat: age of the policyholder where 0 = 21 and younger, 2 = 22-25, 3 = 26-35, 4 = 36-45, 5 = 46-55, 6 = 56-65, 7 = 66 and older
- VehicleAgeCat: age of the vehicle where 2 = 2 years and newer, 3 = 3-5, 4 = 6-10, 5 = 11-15, 6 = 16 years and older

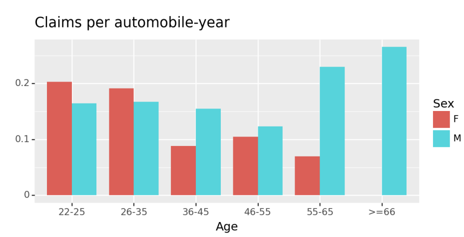
[*a*] Code in python to estimate the claims frequency for automobiles, the expected number of claims per automobile per year, by sex (male and female) and age category (2, 3, 4, 5, 6).

Note that we only focus on automobiles (not other vehicle types), and also exclude records where the sex of the driver is not available.

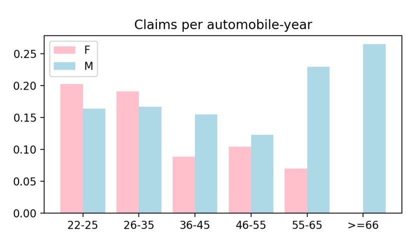
For example, the claims frequency for age category 2 (22-25) for male is estimated as the total number of claims for this group divided by the total amount of automobile-year covered.

① Let your code report(print out) the estimated claims frequency for each combination of sex and age category in a table / dataframe.

② Use `plotnine` to visualize ① similar to the the following



③ Use `matplotlib` to visualize ① similar to the the following



[*b*] Fit a Poisson regression to model `Claim` to the full dataset, using the following features without interactions

- `DiscountCat` where

$$\text{DiscountCat} = \begin{cases} 0 & \text{for} \quad \text{Discount} = 0 \\ 1 & \text{for} \quad \text{Discount} \in \{10, 20, 30\} \\ 2 & \text{for} \quad \text{Discount} \in \{40, 50\} \leftarrow \text{reference level} \end{cases}$$

  and use level 2 as the reference level in the regression.

- `VehicleAgeCat` and use level 6 as the reference level in the regression.

① Let your code report the fit via a summary table which should include the fitted coefficients.

② A Poisson model of `Claim` assumes equal mean and variance, which is not exactly the case for this dataset. Fortunately, we can still use the maximum likelihood estimates of the regression coefficients. One way to correct this over-dispersion is to estimate a multiplier $\hat{\phi}$ and adjust the variance as

$$\hat{\text{Var}}(Y_i) = \hat{\phi}\hat{E}(Y_i)$$

where Y_i is the claims frequency for the i-th policy. The multiplier $\hat{\phi}$ is estimated as

$$\hat{\phi} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\hat{y}_i}$$

Let your code calculate and report $\hat{\phi}$