Problem 1

Download the data and extract records for CSI300 index. After changing the trade_date column to monthly period data type, retain the last record in each month using `drop_duplicates` function, and then calculate the monthly return using `pct_change` function. Finally summarize the statistics. Here we include mean, std, skewness, kurtosis as required, and SW Statistics, SW p-value for further discussion. Note that here we use the Fisher Definition when calculating the Kurtosis, i.e.,

$$Kurtosis = \frac{\mu_4}{\sigma^4} - 3$$

Shown as below:

Table 1: Summary Statistics for CSI300 Monthly Return

Mean	Std	Skewness	Kurtosis	SW Statistics	SW p-value
0.009042	0.081745	0.018308	1.393663	0.971284	0.000228

Then we draw the histogram. Attach the according normal distribution curve using mean and standard deviation. Shown as below:

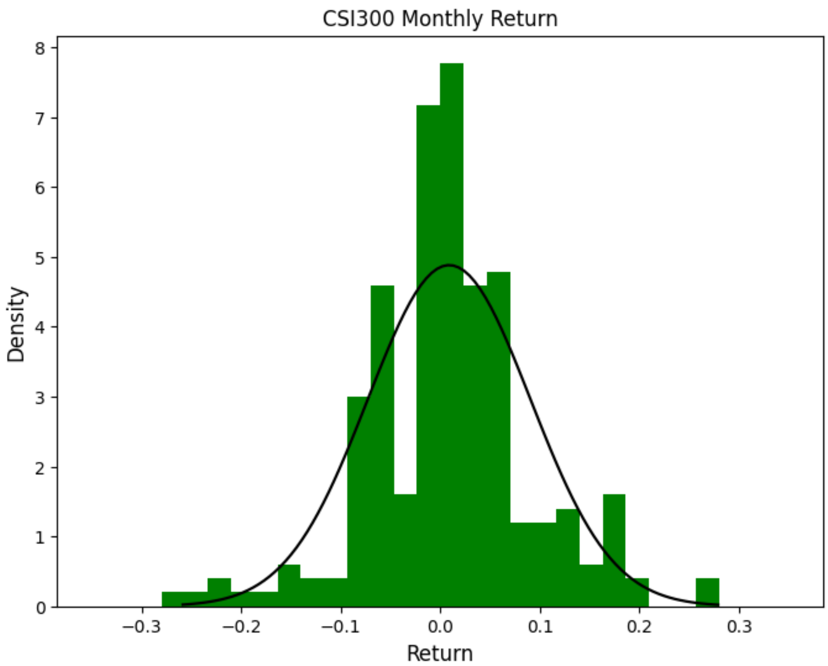

Figure 1: CSI300 Monthly Return

Now we discuss whether returns of the CSI300 index follow a normal distribution.

1. On a intuitive perspective, we directly observe that in Figure 1, the overall shape of the histogram does not match the corresponding normal distribution curve very well.

2. Notice that in Table 1, the Skewness and Kurtosis are greater than zero. This indicates that the distribution of the returns is left-skewed and has a sharper peak than normal distribution.

3. On a rigorous perspective, we conduct Shapiro-Wilk test. Since the p-value < 0.05, we reject H_0 which assumes the returns follow normal distribution.

Therefore, we conclude the returns of the CSI300 index does not follow a normal distribution.

Problem 2

All the data are downloaded as required. Then filter out records that do not belong to A-share mainboard stocks. To calculate weekly market return, we take the average of return at each single week, result shown as below:

```
        date  mkt_return
0    2017-01    0.018377
1    2017-02   -0.032801
2    2017-03   -0.015996
3    2017-04    0.018515
4    2017-05   -0.002261
..       ...         ...
303  2022-49    0.028340
304  2022-50    0.003581
305  2022-51   -0.014400
306  2022-52   -0.047955
307  2022-53    0.012777

[308 rows x 2 columns]
```

Figure 2: Weekly Market Return

Now we begin to test CAPM following section 4 in Chen et al. (2019). Divide the sample period into 3 sub-periods.

1. **1st-period**. For each stock, we regress its excess return on market premium across 2017-2018. Then we get β_i for each stock i.

2. **2nd-period**. To offset idiosyncratic risk while maintaining market premium, we divide stocks into 10 portfolios by sorting β_i obtained from 1st-period using `qcut` function. Then we calculate portfolio return by taking average of return of stocks that belong to each portfolio at each single week. Then for each portfolio, we regress its excess return on market premium across 2019-2020. Thus we replicate Table 2 in Chen et al. (2019). Shown as below:

Table 2: Time Series Regression Results of the Second Period of Stock Portfolio

Portfolio	α	α_t	α_p	β	β_t	β_p	R-Squared
1	-0.00120	-1.706	0.0910	0.77932	34.705	0.0	0.923
2	0.00021	0.356	0.7224	0.87666	46.376	0.0	0.956
3	0.00011	0.278	0.7815	0.91094	70.605	0.0	0.980
4	0.00096	2.186	0.0311	0.93897	66.730	0.0	0.978
5	0.00001	0.019	0.9852	1.00924	88.912	0.0	0.988
6	0.00029	0.890	0.3758	1.04444	98.947	0.0	0.990
7	-0.00012	-0.276	0.7834	1.04772	72.568	0.0	0.981
8	0.00002	0.048	0.9621	1.09918	73.235	0.0	0.982
9	0.00031	0.550	0.5836	1.13993	62.694	0.0	0.975
10	-0.00060	-0.944	0.3475	1.15360	57.007	0.0	0.970

From table 2, we discover that:

- The β of each portfolio is around 1 with basically small β_p value which indicates that stock return are greatly affected by market return.
- The α of each portfolio is approaching 0 and 90% of α_p exceeds 0.05. Thus we lack sufficient evidence to reject H_0 that alpha equals zero.

3. **3rd-period**. We calculate mean excess return by taking the average of excess return for each portfolio across 2021-2022. For each portfolio, we regress its mean excess return on β obtained from 2nd-period. Result shown as below:

Table 3: Cross-Sectional Regression Results of the Third Period of Stock Portfolio

	γ_0	γ_1	R-Squared	F-statistics	P-value
Coefficient	-0.0016	0.0031	0.563	10.32	0.012
t-value	-1.617	3.213			

From table 3, we discover that γ_1 is significantly positive, which validates the positive relation between excess return and systematic risk.