Contents

Introduction

This report analyzes weekly return data from a selection of Exchange Traded Funds (ETFs) using Python. The analysis is divided into a descriptive part and a statistical inference part. The goal is to evaluate data quality, understand return distributions, and make portfolio optimization decisions based on statistical measures.

Descriptive Analysis

a) Data overview

The dataset used in this analysis consists of weekly returns for 95 different Exchange Traded Funds (ETFs). The dataset spans a period from May 5, 2006, to May 8, 2015, covering a total of 454 weekly observations. Each observation represents the weekly return, calculated as the ratio between the closing and opening prices for that week minus one (i.e. $Return = \frac{P_{end}}{P_{start}} - 1$.

For this project, we focus on four selected ETFs:

- **AGG** – iShares Core U.S. Aggregate Bond ETF
- **VAW** – Vanguard Materials ETF
- **IWN** – iShares Russell 2000 Value ETF
- **SPY** – SPDR S&P 500 ETF Trust

These ETFs were chosen to reflect a variety of asset classes including bonds, large-cap U.S. stocks, small-cap value stocks, and sector-focused equities.

An initial inspection of the data confirms that there are no missing values for the selected ETFs during the period. The data is clean and suitable for further statistical analysis.

b) Summary Statistics

The table below presents summary statistics for the weekly returns of the selected ETFs (AGG, VAW, IWN, and SPY) over 454 observations from 2006 to 2015.

AGG, a bond ETF, shows the lowest average return (0.0266%) and the lowest volatility (standard deviation ≈ 0.00598). VAW has the highest average return (0.1794%) but also the highest variance and standard deviation, indicating more risk.

All ETFs display a relatively wide interquartile range (Q3 - Q1), with VAW and IWN having the most dispersed return distributions. The median returns are generally close to the mean, though some skewness is expected and will be examined in the next section.

	Number of obs	Sample mean	Sample variance	std. dev.	Lower quartile	Median	Upper quartile
AGG	454	0.000266	0.000036	0.005976	-0.002973	0.000237	0.003893
VAW	454	0.001794	0.001302	0.036083	-0.016096	0.004798	0.019685
IWN	454	0.001188	0.001025	0.032015	-0.014305	0.003120	0.019056
SPY	454	0.001360	0.000614	0.024786	-0.011325	0.004216	0.014498

Table 1: Summary statistics of weekly returns for the selected ETFs from 2006 to 2015

Column	Meaning
Number obs.	Number of weekly observations (454 for all = no missing values)
Sample mean	Average weekly return. VAW has the highest (~0.18%), AGG is lowest.
Sample Variance	Variance of returns (risk): higher = more volatile
Std. dev.	Standard deviation (square root of variance): also shows volatility
Lower quartile	25th percentile - lower bound of the middle 50%
Median	50th percentile - the midpoint of the distribution
Upper quartile	75th percentile - upper bound of the middle 50%

c) Graphical Analysis

To visualize the distribution of weekly returns, a combined empirical density plot and a boxplot were created (below) for the selected ETFs (AGG, VAW, IWN, SPY). The density plot (Figure 1) shows how returns are spread across values, while the boxplot (Figure 2) highlights the central tendency and outliers.

AGG shows a narrow, concentrated distribution, confirming its low volatility. VAW and IWN show wider spreads and several outliers, consistent with their higher standard deviations. SPY appears balanced and symmetrical, indicating moderate volatility.

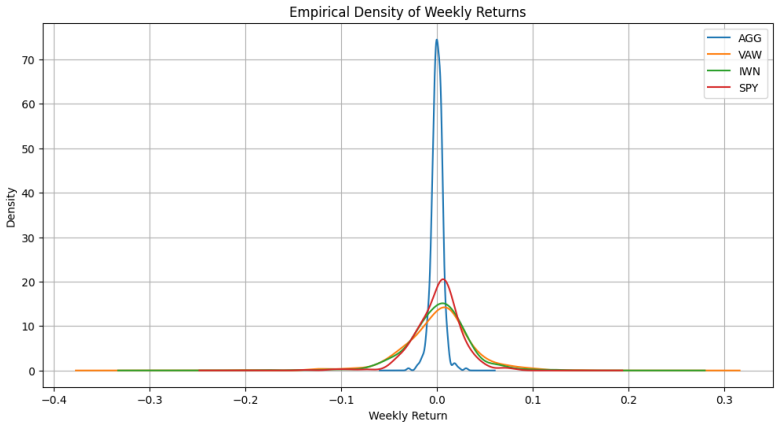


Figure 1: Empirical Density of Weekly Returns

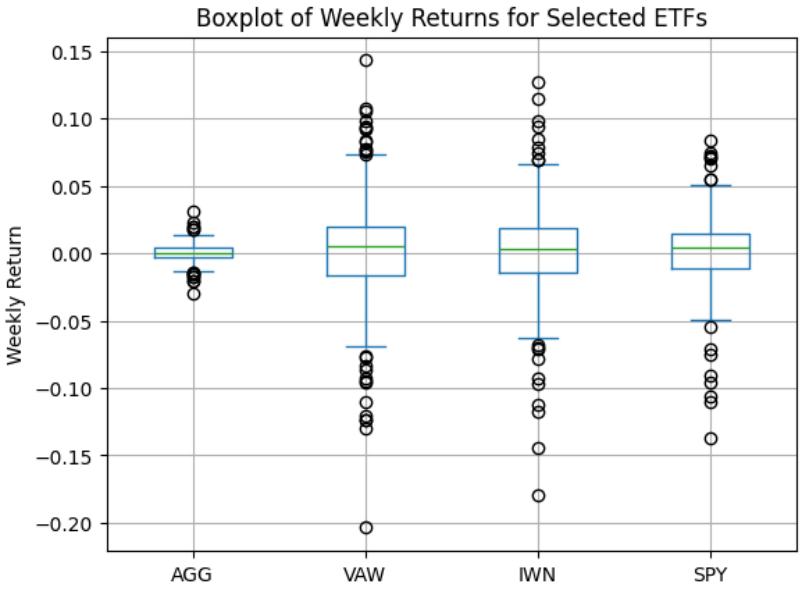


Figure 2: Boxplot of Weekly Returns for Selected ETFs

Distribution Description

Based on the density plots and boxplots, the distribution of weekly returns varies notably between the ETFs.

- AGG displays a narrow, symmetric distribution centered around zero, indicating low volatility. Its boxplot confirms this with a compact interquartile range and very few outliers. Weekly returns for AGG range approximately from -2.96% to $+3.05\%$, which is consistent with the low-risk profile of bond investments.
- VAW shows a wider distribution with a slight right skew, meaning that while most returns are low or near zero, there are some high positive outliers. The boxplot shows several outliers on the high side. Weekly returns range from -20.37% to $+14.30\%$, indicating high volatility.
- IWN also exhibits a relatively wide and slightly right-skewed distribution, like VAW. It has a large number of outliers and a widespread in the boxplot. The weekly return range spans from -17.97% to $+12.67\%$, which is typical for small-cap value ETFs known for higher risk.
- SPY, which tracks the S&P 500 index, has a slightly right-skewed distribution with more frequent large positive returns. The density is centered around 0.0014, but the histogram and QQ plot show positive skewed and heavier upper tails. The boxplot reflects this asymmetry with a longer upper whisker and mild outliers. Weekly returns range from –13.76% to +8.33%, indicating moderate volatility typical for a broad equity index fund.

In summary, AGG is the most stable and symmetric, while VAW and IWN are more volatile and right skewed with more extreme values. SPY sits in the middle in terms of volatility and shape.

Problem 1 – Portfolio Construction

d) Covariance and Correlation

Table 3: Covariance Matrix of Weekly Returns

	AGG	VAW	IWN
AGG	0.000036	-0.000043	-0.000026
VAW	-0.000043	0.001302	0.000984
IWN	-0.000026	0.000984	0.001025

Table 4: Correlation Matrix of Weekly Returns

	AGG	VAW	IWN
AGG	1.000000	-0.197568	-0.135262
VAW	-0.197568	1.000000	0.851641
IWN	-0.135262	0.851641	1.000000

To analyze the co-movement between AGG, VAW, and IWN, we calculated their weekly return covariance and correlation matrices.

Table 3 shows that AGG has a positive variance of 0.000036, indicating relatively low but non-zero variability in its weekly returns. However, AGG's covariance with VAW (-0.000043) and with IWN (-0.000026) is slightly negative, indicating a weak inverse relationship with those equity ETFs.

Table 4 shows the correlation matrix, where AGG has a:

- -0.198 correlation with VAW
- -0.135 correlation with IWN

This suggests that while AGG moves mostly independently, there is a weak tendency for it to move in the opposite direction of these two stock-based ETFs - useful for portfolio diversification.

By contrast, VAW and IWN have a strong positive correlation (≈ 0.85), showing similar movement patterns, as expected from two equity-focused ETFs.

e) Pairwise Scatterplots

Figure 3-6 shows scatterplots of weekly returns for all combinations of the four ETFs: AGG, VAW, IWN, and SPY. The strongest linear relationship appears between VAW and IWN, indicating a high positive correlation, consistent with the correlation matrix (~0.85).

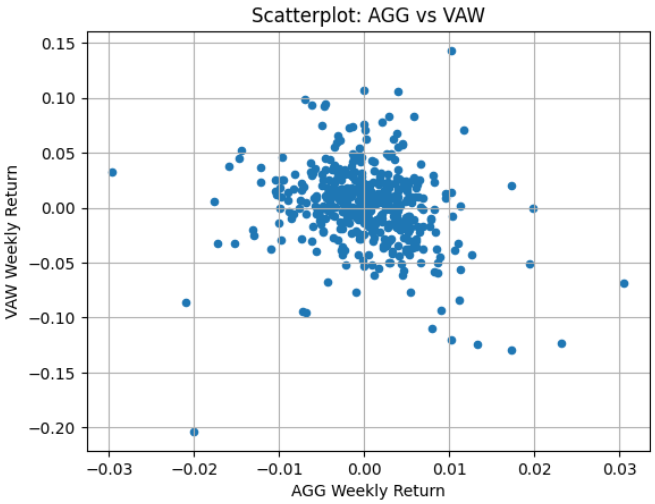


Figure 4: AGG vs VAW

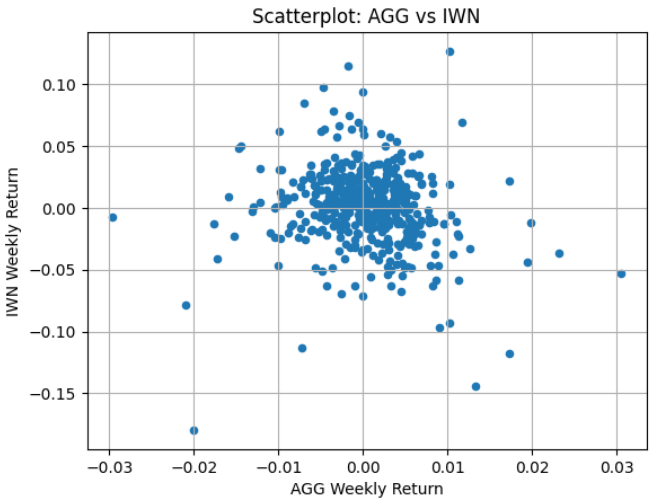


Figure 3: AGG vs IWN

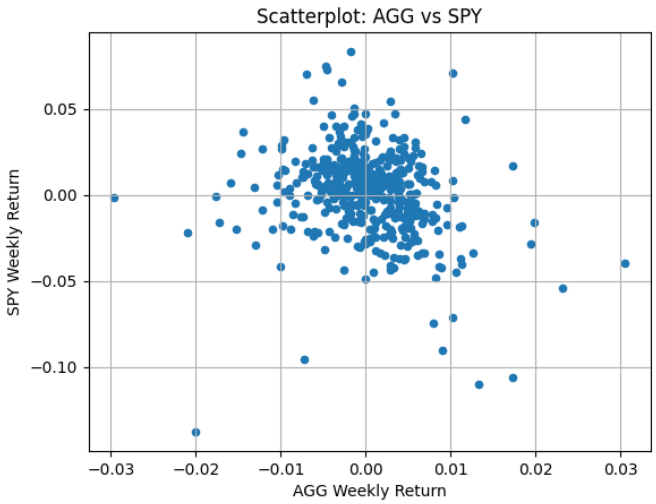


Figure 5: AGG vs SPY

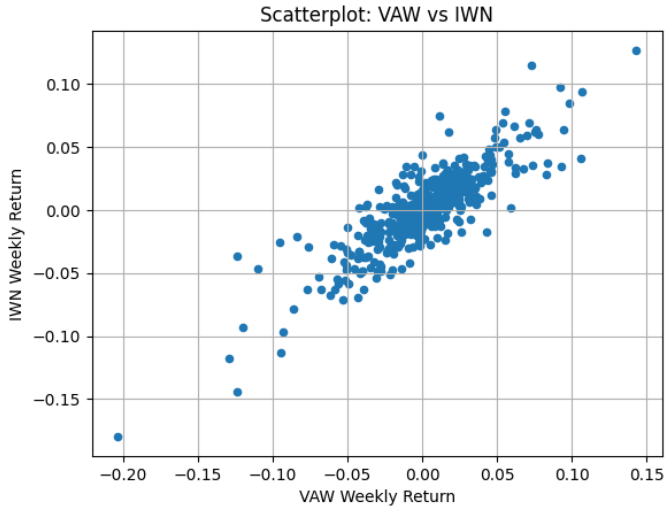


Figure 6: VAW vs IWN

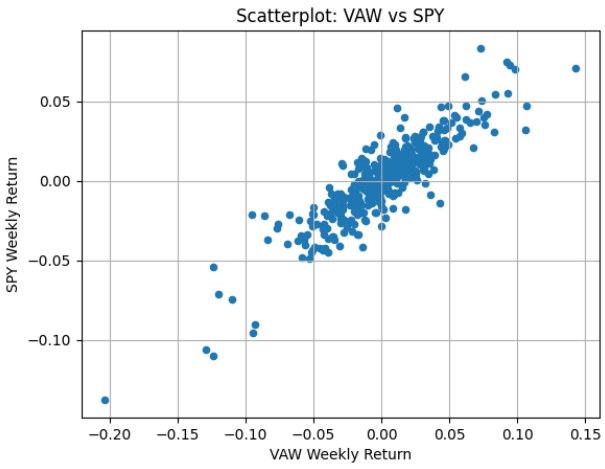


Figure 7: VAW vs SPY

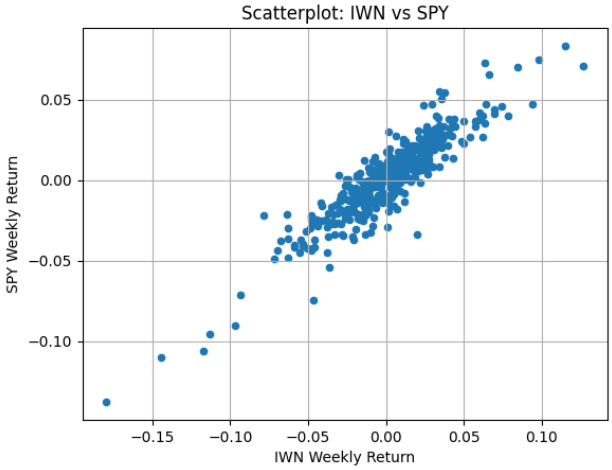


Figure 8: IWN vs SPY

Portfolio Variance Minimization

f1)

To construct a two-ETF portfolio, we define a random variable P_1 representing the weekly return of a portfolio composed of AGG and VAW:

$$P_1 = \alpha \cdot X_{AGG} + (1 - \alpha) \cdot X_{VAW}$$

Here, α is the share of the investment allocated to AGG, and $1 - \alpha$ is allocated to VAW. The goal is to find the value of α that minimizes the portfolio variance $Var(P_1)$, providing the optimal risk diversification between the two assets.

To generalize this approach, we define similar expressions for all combinations of the four selected ETFs (AGG, VAW, IWN, and SPY). Each portfolio return P_i is modeled as:

$$P_1 = \alpha \cdot X_{AGG} + (1 - \alpha) \cdot X_{VAW}$$

$$P_2 = \alpha \cdot X_{AGG} + (1 - \alpha) \cdot X_{IWN}$$

$$P_3 = \alpha \cdot X_{AGG} + (1 - \alpha) \cdot X_{SPY}$$

$$P_4 = \alpha \cdot X_{VAW} + (1 - \alpha) \cdot X_{VAW}$$

$$P_5 = \alpha \cdot X_{VAW} + (1 - \alpha) \cdot X_{SPY}$$

$$P_6 = \alpha \cdot X_{IWN} + (1 - \alpha) \cdot X_{SPY}$$

These combinations will be evaluated to determine which portfolio provides the lowest variance and best tradeoff between risk and return.

f2)

To express the portfolio variance mathematically, we use the formula:

$$Var(P_1) = \alpha^2 \cdot Var(X_{AGG}) + (1 - \alpha)^2 \cdot Var(X_{VAW}) + 2\alpha(1 - \alpha) \cdot Cov(X_{AGG}, X_{VAW})$$

By inserting the values calculated earlier:

$$Var(P_1) = \alpha^2 \cdot 0.000036 + (1 - \alpha)^2 \cdot 0.001302 + 2\alpha(1 - \alpha) \cdot (-0.000043)$$

This function describes how the portfolio variance depends on the weight α\alphaα invested in AGG. In the next step, we will visualize this function and determine the value of α\alphaα that minimizes the variance.

f4)

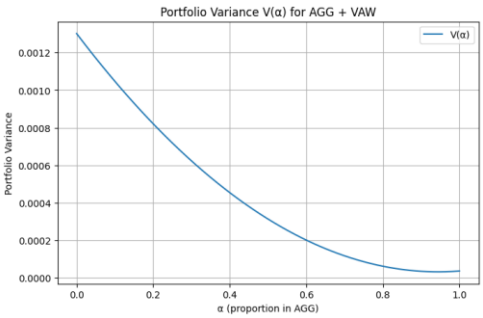


Figure 9

A graph of the portfolio variance function $V(\alpha)$) was created for the ETF pair AGG and VAW. As shown in Figure 9, the function is convex and reaches a clear minimum. This visual representation confirms that there is an optimal weight $\alpha \in [0,1]$, that minimizes risk.

The minimum portfolio variance is obtained at:

- $\alpha_m = 0.9439$
- $Var(P_1) = 0.000032$

This result implies that the lowest-risk portfolio between AGG and VAW is achieved by investing 94.39% in AGG and 5.61% in VAW.

f5)

The value $\alpha_m = 0.9439$ lies within the interval $[0,1]$, indicating a feasible portfolio allocation under standard investment rules (no short selling).

Had $\alpha_m < 0$, it would imply short-selling AGG, and if $\alpha_m > 1$, it would imply short-selling VAW. These are not permitted in this model, so we only consider allocations within the feasible region.

The fact that the optimal point lies within the interval confirms that a valid minimum-variance portfolio can be constructed with this ETF pair.

f6)

The table below presents the optimal portfolio weights α_m that minimize variance for each ETF pair. Each α_m indicates how much of the portfolio should be allocated to the first ETF in the pair to minimize portfolio risk:

	Portfolio	α_m (in AGG)	Min Variance	α_m (in VAW)	α_m (in IWN)
0	AGG + VAW	0.9459	0.000031	NaN	NaN
1	AGG + IWN	0.9439	0.000032	NaN	NaN
2	AGG + SPY	0.9038	0.000029	NaN	NaN
3	VAW + IWN	NaN	0.001020	0.1142	NaN
4	VAW + SPY	NaN	0.000614	0.0000	NaN
5	IWN + SPY	NaN	0.000614	NaN	0.0

Table 5: Optimal Portfolio Weights and Minimum Variances

These results show that the lowest-risk portfolios are consistently achieved when allocating most of the investment to AGG. Portfolios composed only of equity ETFs (VAW, IWN, SPY) result in significantly higher variance, unless one ETF is completely excluded.

f7)

The table below presents the minimum portfolio variance achieved for each ETF combination. These values reflect the lowest possible risk attainable through optimal allocation between the two ETFs:

Table 6: Minimum Variance by ETF Pair

Portfolio	Minimum Variance
AGG + VAW	0.000031
AGG + IWN	0.000032
AGG + SPY	0.000029
VAW + IWN	0.001020
VAW + SPY	0.000614
IWN + SPY	0.000614

As expected, portfolios that include AGG (a bond ETF) consistently have much lower risk compared to combinations of two equity-based ETFs. The AGG + SPY combination yields the lowest minimum variance, suggesting the best risk-reducing potential among all pairs.

f8)

The table below shows the expected weekly return $E(P_i)$ for each portfolio when invested optimally according to the minimum variance weights α_m:

Table 9: Expected Weekly Returns of Minimum-Variance Portfolios

	Portfolio	α_m (in AGG)	$E(P_i)$	α_m (in VAW)	α_m (in IWN)
0	AGG + VAW	0.9459	0.000348	NaN	NaN
1	AGG + IWN	0.9439	0.000317	NaN	NaN
2	AGG + SPY	0.9038	0.000371	NaN	NaN
3	VAW + IWN	NaN	0.001257	0.1142	NaN
4	VAW + SPY	NaN	0.001360	0.0000	NaN
5	IWN + SPY	NaN	0.001360	NaN	0.0

These results show that portfolios involving only equity ETFs (e.g., SPY, IWN, VAW) have higher expected returns, but as shown earlier, they also come with higher variance. In contrast, portfolios involving AGG have significantly lower risk but also more modest returns.

f9)

Among the six ETF combinations analyzed, we select the AGG + SPY portfolio as the optimal choice.

This portfolio achieves the lowest portfolio variance of all combinations (0.0000290.0000290.000029) while also offering the highest expected return (0.0003710.0003710.000371) among the low-risk portfolios. It balances risk and returns effectively by combining a broad U.S. bond index (AGG) with a well-diversified large-cap U.S. equity index (SPY).

In contrast, portfolios consisting only of equities (e.g., VAW + SPY or IWN + SPY) offer higher expected returns but come with significantly greater variance and less diversification benefits.

Therefore, the AGG + SPY portfolio is chosen for its **optimal tradeoff between return and risk**, making it the most efficient portfolio in this context.

Problem 2- Best investment

g)

To evaluate whether it's more beneficial to invest in ETFs or save money without return, we model the weekly returns of the four selected ETFs (AGG, VAW, IWN, SPY) as random variables:

$$X_i \sim N(\mu_i, \sigma_i^2)$$

where μ_i and σ_i^2 represent the expected value and variance of weekly returns for ETF iii, respectively.

Assumptions:

- Returns are independent and identically distributed
- Returns follow a normal distribution

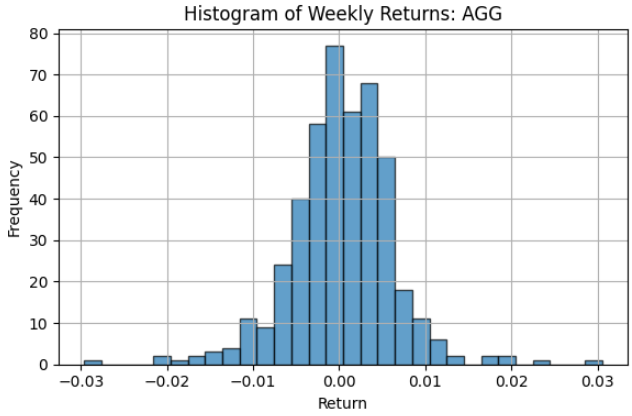
- No structural breaks or time-dependent effects in the data

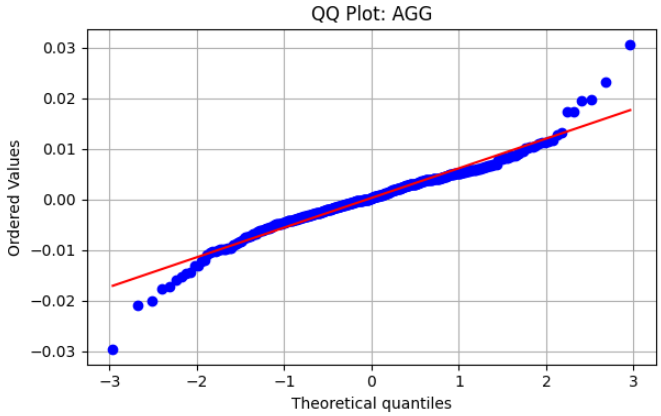
We estimate the parameters μ_i and σ_i^2 from historical weekly return data.

Model Validation

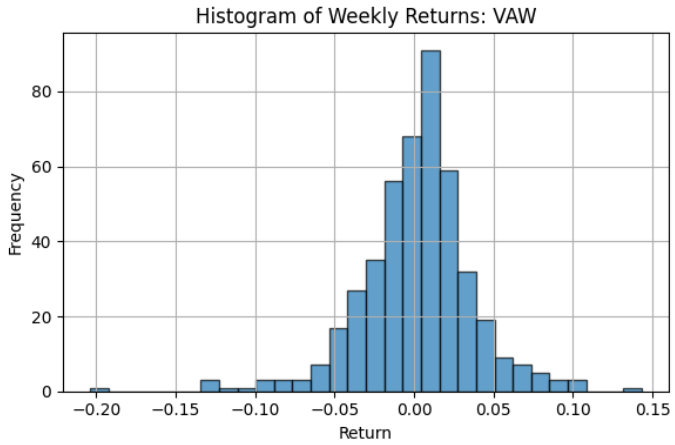
To assess whether the assumption of normally distributed returns is valid for the selected ETFs, we examined both histograms and QQ plots of the weekly returns. The results are summarized below:

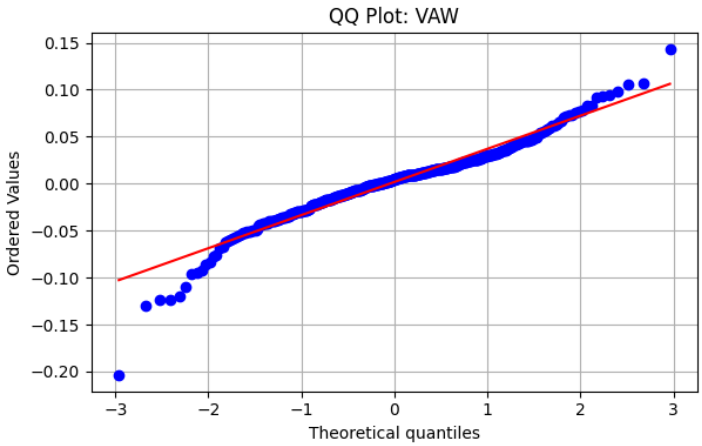
AGG: The distribution of AGG returns is approximately symmetric and bell-shaped. The QQ plot shows minor deviations at the tails but closely follows the diagonal line, indicating that the normality assumption is reasonable for AGG. This aligns with its low volatility and stable return profile as a bond ETF.



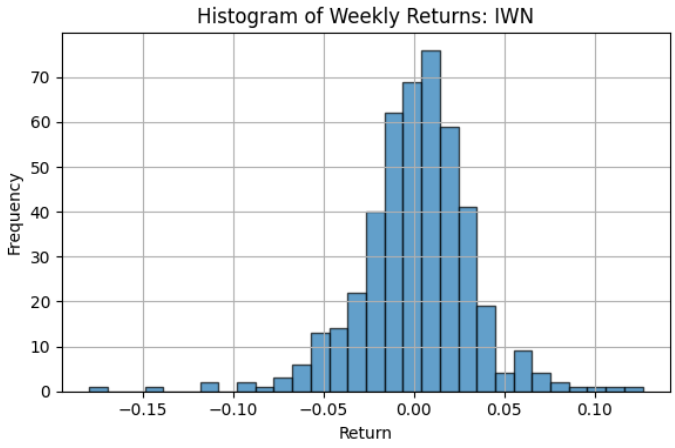


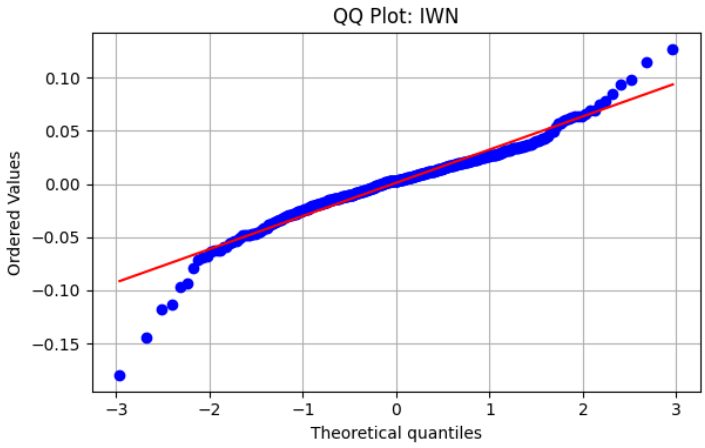
VAW: The distribution shows noticeable right skewness and heavy tails. The QQ plot deviates significantly from normality, especially in the extremes. Normality is not fully satisfied.



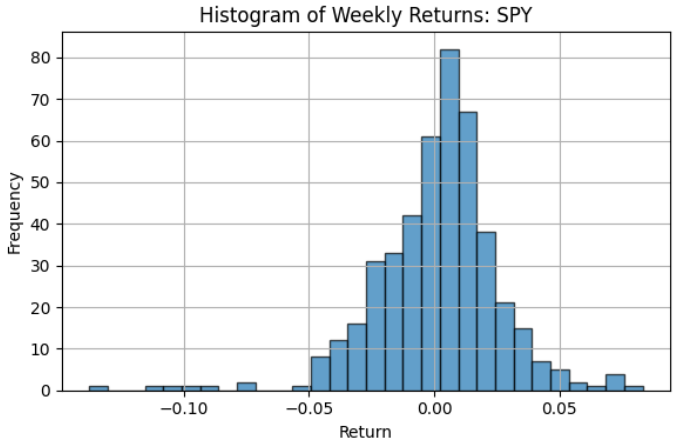


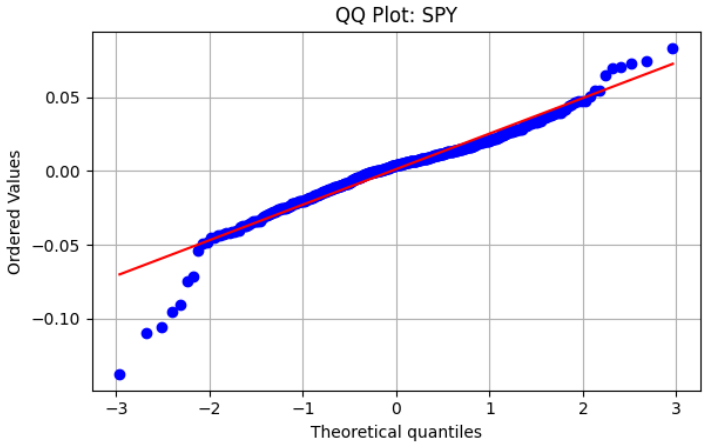
IWN: The distribution of IWN returns is right skewed with heavy tails, as seen in both the histogram and the QQ plot. The deviations from the normal line are more pronounced at the extremes, indicating that the normality assumption is not fully satisfied.





SPY: The distribution of SPY returns is slightly right-skewed, as seen in the histogram with a longer upper tail and confirmed by the QQ plot, where the data deviates above the line at the right end. This suggests that while most weekly returns are centered around the mean, there are more extreme positive returns than expected under a normal distribution. Nevertheless, the central portion of the distribution follows the normal pattern fairly well, making the normality assumption reasonably acceptable for inference.





Due to the large sample size ($n = 454$), we invoke the Central Limit Theorem (Theorem 3.14), which states that the distribution of the sample mean will tend toward a normal distribution even when the underlying data is not perfectly normal. This supports the validity of using normal-based inference for the expected returns μ_i.

h) Confidence Intervals for the Mean Weekly Return

To assess the uncertainty around the estimated weekly return for each ETF, we calculated 95% confidence intervals using the formula:

$$\mathrm{CI}_{95\%} = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}} \quad , \quad z = 1.96$$

Where $\bar{x}$ is the sample mean, s is the sample standard deviation, and $n = 454$ is the number of observations. The results are shown in Table 10:

Table 10: 95% Confidence Intervals for Weekly Returns

	ETF	Mean	95% CI Lower	95% CI Upper	CI Width
0	AGG	0.000266	-0.000284	0.000815	0.001099
1	VAW	0.001794	-0.001525	0.005113	0.006638
2	IWN	0.001188	-0.001757	0.004133	0.005890
3	SPY	0.001360	-0.000920	0.003640	0.004560

The confidence intervals vary in width. AGG has the narrowest interval, while VAW and IWN have wider intervals due to their higher return volatility (standard deviation). This confirms that higher volatility results in more uncertainty when estimating the mean.

Conclusion

Based on the statistical analysis of weekly return data for the ETFs AGG, VAW, IWN, and SPY, we conclude that investing in ETFs offers a significantly better long-term outcome than saving money under the pillow, which yields no return and no growth.

All four ETFs have positive average returns, with expected weekly returns ranging from 0.027% (AGG) to 0.179% (VAW). Even the most conservative ETF (AGG) outperforms non-investment when considering the long term.

Though equity ETFs such as VAW and IWN have higher volatility and wider confidence intervals, their potential for greater return makes them attractive for risk-tolerant investors. On the other hand, AGG and SPY offer a good balance between risk and return, especially SPY, which has moderate volatility and relatively high expected return.

In conclusion, investing in a diversified ETF portfolio provides a clear statistical advantage over not investing, especially when long-term compounding and inflation protection are considered.

Appendix

The complete Python code, data analysis, and visualizations used in this project are included in the Jupyter Notebook file submitted separately with the report. All tables and figures referenced in this document were generated from this notebook.