FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**Min Wang**
Team member 2	**DIAWARA Louckmane**
Team member 3	**Ashish Kumar**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
N/A

1. On top left of your screen click on File →Download → Microsoft Word (.docx) to download this template
2. **Upload the template in Google Drive and share it with your group members**
3. **Delete this page with the requirements before submitting your report. Leaving them will result in an increased similarity score on Turnitin.**

Keep in mind the following:

- Make sure you address all the questions in the GWP assignment document published in the Course Overview.
- Follow the "Submission requirements and format' instructions included in each Group Work Project Assignment, including report length.
- **Including in-text citations and related references is mandatory for all submissions.** You will receive a '0' grade for missing in-text citations and references, or penalties for partial completion. Use the **In-Text Citations and References Guide** to learn how to include them.
- Additional writing aids: Anti-Plagiarism Guide, Academic Writing Guide, Online Writing Resources.
- To avoid an increase in the Turnitin similarity score, **DO NOT copy the questions** from the GWP assignment document.
- Submission format tips:
    - **Use the same font type and size and same format throughout your report.** You can use Calibri 11, Arial 10, or Times 11.
    - Do NOT split charts, graphs, and tables between two separate pages.
    - **Always include the axes labels and scales in your graphs as well as an explanation of how the data should be read.**
- Use the LIRN Library for your research. It can be accessed via the left navigation pane inside the WQU learning platform.
- Carefully read **Academic Policy on the use of AI** explaining how the use of AI tools is restricted and regulated. Severe penalties apply for excessive and improper use of AI

The PDF file with your report must be uploaded separately from the zipped folder that includes any other types of files. This allows Turnitin to generate a similarity report.

Solutions

1. Data Quality

1(a) – Example of Poor-Quality Structured Data

No	Field / Column	Example of Bad Data	Nature of Problem	Real-World Consequence
1	Customer ID	`001A`	Inconsistent pattern (alphanumeric)	Can't join with numeric IDs in other tables
2	Name	`john d`, `JOHN DOE`, `John D.`	Inconsistent capitalization	Duplicate-detection fails
3	Date of Birth	`13/31/1990`	Invalid date format	Parsing errors in SQL/Python
4	Age	`twenty five`	Wrong data type (text vs numeric)	Aggregations return NaN
5	Gender	`M`, `male`, `Male`, `man`	Non-standardized categories	Gender-based analysis inaccurate
6	Country	`IND`, `India`, `Bharat`	Mixed codes and names	Lookup tables don't match
7	Account Balance	`₹ 1 20 000`, `120000INR`	Wrong numeric format	Currency conversion fails
8	Interest Rate	`0.05%` vs `5%`	Scale and symbol inconsistency	Misleading APR calculations
9	Transaction Date	`2024/31/03`	Reversed day/month	Wrong chronological order
10	Transaction Amount	`-5000` vs `5000`	Negative sign used incorrectly	Misrepresents cash flow
11	Credit Score	`850+`, `good`, `NA`	Mixed numeric and text	Impossible to model quantitatively
12	Phone Number	`12345` or `+91-XXXXXXXXX`	Length variations	SMS / OTP delivery fails

13	Email ID	`john[at]mail.com`	Invalid syntax	Message bounce-backs
14	PAN Number	Missing or duplicated	Uniqueness violation	Regulatory non-compliance
15	Loan Type	`Home`, `Mortgage`, `HOUSE`	Non-standard labels	Aggregation errors
16	Collateral Value	`ten lakh`	Textual entry instead of number	Under-/over-valuation
17	Branch Code	`null` or `9999` (dummy)	Missing data	Loss of traceability
18	Currency	`INR`, ₹, `rupees`	Mixed representation	FX conversion issues
19	Account Status	`Actv`, `Active Account`, `A`	Abbreviations inconsistent	Reporting errors
20	Closing Date	Blank fields	Missing information	Lifecycle analysis impossible
21	Opening Balance	`0.000000e+00`	Excessive decimal precision	Confuses non-technical users
22	Debit/Credit Flag	`Dr`, `debit`, `1`	Non-standard encoding	Aggregation logic fails
23	Branch Name	`Mumbai East (Old)` vs `Mumbai-E`	Non-uniform naming	Geo mapping fails
24	Risk Category	`Low`, `Medium`, `3`	Mixed ordinal and numeric scale	Statistical bias
25	Portfolio Code	`PF-1`, `001`, `A1`	Code format not standardized	Sorting and indexing issues
26	Transaction Time	`25:61:00`	Impossible time value	Log processing crashes
27	Net Asset Value	`12,34,56.78`	Wrong comma placement	Parser fails to read float

28	Security Ticker	`rel`, `RELIANCE`, `RLNC`	Alias mismatch	Market data link breaks
29	ISIN Code	Typo or missing check digit	Integrity error	Wrong security mapping
30	Dividends Paid	`None` in some rows	Null values ignored in sum	Understated returns
31	Fund Manager ID	Duplicate IDs	Non-unique primary key	Attribution errors
32	Expense Ratio	5 (vs 0.05)	Scale misinterpretation	Overstated costs by ×100
33	Rating Agency	`S&P`, `SP`, `Standard&Poor`	Naming variations	Text join fails
34	Bond Maturity Date	`2023-00-15`	Invalid month	Query error
35	Duration	`--` or `n/a`	Placeholder text	Numeric ops fail
36	Modified Duration	`0.00E+00`	Scientific notation misread	Misleading visualization
37	Yield to Maturity	Blank or negative	Logical impossibility	Analyst confusion
38	Repo Rate	`6,25`	European comma decimal	Wrong float parse
39	Inflation Rate	`7%` and `0.07` mixed	Unit inconsistency	Misleading charts
40	Sector	`Tech.`, `Technology`, `IT`	Label inconsistency	Aggregation errors
41	Market Cap	`₹ ten crore`	Text input	Fails in numeric sort
42	Return on Equity	`>20%`	Symbolic value	Can’t convert to float

43	Asset Class	`Equity/Bonds` mixed	Multiple entries in one cell	Parsing problem
44	Benchmark	`Nifty50 + Sensex`	Multi-benchmarks	Ambiguous reference
45	File Encoding	UTF-8 vs Latin-1	Encoding mismatch	Data corruption
46	Decimal Separator	`.` vs `,`	Regional format conflict	Inconsistent float read
47	Header Row	Missing or misaligned	Structural error	Mis-labeled columns
48	Duplicate Rows	Entire record repeated	Redundant entries	Skews averages
49	Outlier Value	`Interest Rate = 99%`	Extreme outlier	Invalid data point
50	Wrong Join Key	`AccountID` typed as `Account_Id`	Schema mismatch	Merge fails

1(b) Explanation – Why the Structured Data Is of Poor Quality

When I look at the dataset above, the problems are immediately visible even before any code is run. Several fields are incomplete, inconsistent, or typed incorrectly—for instance, numeric values appear as text, and dates follow different regional formats. Some entries use symbols like “–” or “N/A” instead of proper nulls, while others show different spellings for the same country or category.

What this tells me is that the data lacks uniformity and validation rules. In finance, this kind of noise can completely distort metrics such as average balance or loan exposure. It also breaks automated joins across tables, meaning even the most accurate models downstream could produce misleading insights. Essentially, the dataset fails the fundamental principles of **accuracy, consistency, and completeness** (Redman 1998), which form the backbone of reliable financial analytics and are also outlined in the **ISO 25012 Data Quality Model** (ISO 2019).

1(c) Example of Poor-Quality Unstructured Data

> *"Client complained loudly about double charge!! said refund issued?? check later – possible technical glitch or fraud – account no maybe 6781 or 6787?? need follow up (urgent)"*

This kind of data often comes from free-form text notes in CRM systems, email logs, or call-center transcripts. It's not stored in rows and columns but in raw narrative form—which makes it far more difficult to clean or analyze without manual judgment (Madnick et al., 2009).

(d) Explanation – Why the Unstructured Data Is of Poor Quality

The problem with this note is that it mixes facts, opinions, and incomplete identifiers in a single paragraph. There's emotional language ("complained loudly"), uncertainty ("maybe 6781 or 6787"), and no clear separation between verified and assumed information. The syntax and abbreviations vary from person to person, so even a good natural-language model might misinterpret intent or sentiment. In financial systems, unstructured text like this can easily lead to wrong classifications—for example, flagging a technical error as a fraud case.

Good-quality unstructured data would capture the same information in a consistent template with standardized tags (e.g., *Issue Type*, *Severity*, *Action Taken*). In short, this snippet fails because it's ambiguous, incomplete, and lacks metadata, which prevents both humans and algorithms from drawing confident conclusions (Wang & Strong 1996).

References

- International Organization for Standardization. *ISO 25012: Systems and Software Engineering — Data Quality Model.* ISO, 2019.
- Madnick, Stuart E., Richard Y. Wang, and Yang W. Lee. *Improving Data Quality for Decision Making: Good Data Makes Better Decisions.* MIT Information Quality Program, 2009.
- Redman, Thomas C. *Data Quality: Management and Technology.* Bantam Books, 1998.
- Wang, Richard Y., and Diane M. Strong. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, vol. 12, no. 4, 1996, pp. 5–34.

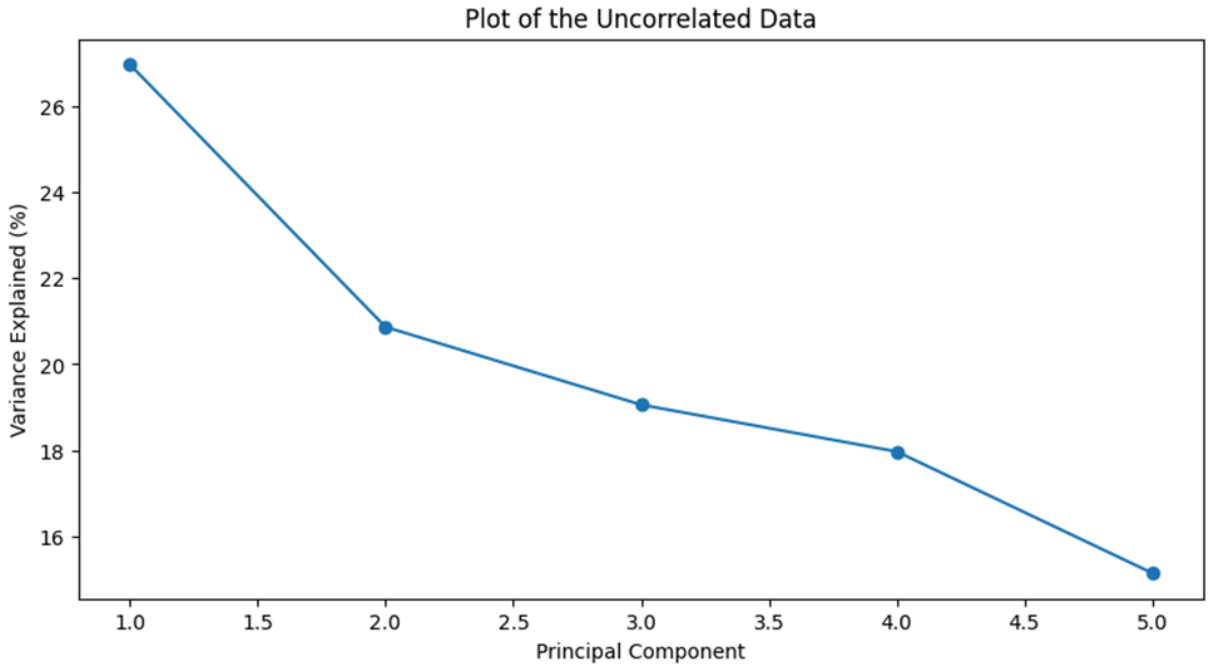
3. Exploiting Correlation

3-a-c) Variance Explanation for Simulated Uncorrelated Data

When PCA (Principal Component Analysis) was performed on the simulated uncorrelated Gaussian yield changes, the variances of the five principal components were relatively evenly distributed, which indicates the variables were uncorrelated. There are five yield changes that were generated independently of one another and that each followed the same normal distribution, meaning no single component can dominate the variance. Component 1 explains 27% of the variance, Component 2 explains 21%, and Component 3 explains 19% of the variance, leading to the conclusion that there is no dominant factor in the data. This is what we would expect from uncorrelated data sets PCA would evenly distribute the variance and there would be no single primary direction of variance (Hastie et al.).

Component	Explained Variance Ratio	Variance Explained (%)
Component 1	0.2698	**26.98%**
Component 2	0.2087	**20.87%**
Component 3	0.1906	**19.06%**
Component 4	0.1797	**17.97%**
Component 5	0.1513	**15.13%**

3d: Screeplot of the Variance Explained for Each Component



A screeplot demonstrates a gradual decline in trend from Principal Components 1 through 5 where explained variance ratios were 0.2698, 0.2087, 0.1906, 0.1797, and 0.1513. The absence of a sharp "elbow" in this gradual decline suggests that no one component captures a large portion of the total variance. This is consistent with the uncorrelated nature of the simulated data.

3h: Comparing the Variances of Individual Components

From the FRED API in Python, we use real yields. We collected the market yield on U.S. Treasury securities at 1-month, 3-month, 6-month, 1-year, and 2-year. The data represent daily closing yields quoted on an investment basis from April 8, 2025, to October 8, 2025 (Federal Reserve Bank of St. Louis).

Date	DGS1MO	DGS3MO	DGS6MO	DGS1	DGS2
2025-04-08	4.36	4.31	4.14	3.83	3.71
2025-04-09	4.36	4.35	4.23	4.03	3.91
2025-04-10	4.36	4.34	4.17	3.97	3.84
2025-04-11	4.37	4.34	4.21	4.04	3.96
2025-04-14	4.34	4.33	4.21	3.99	3.84

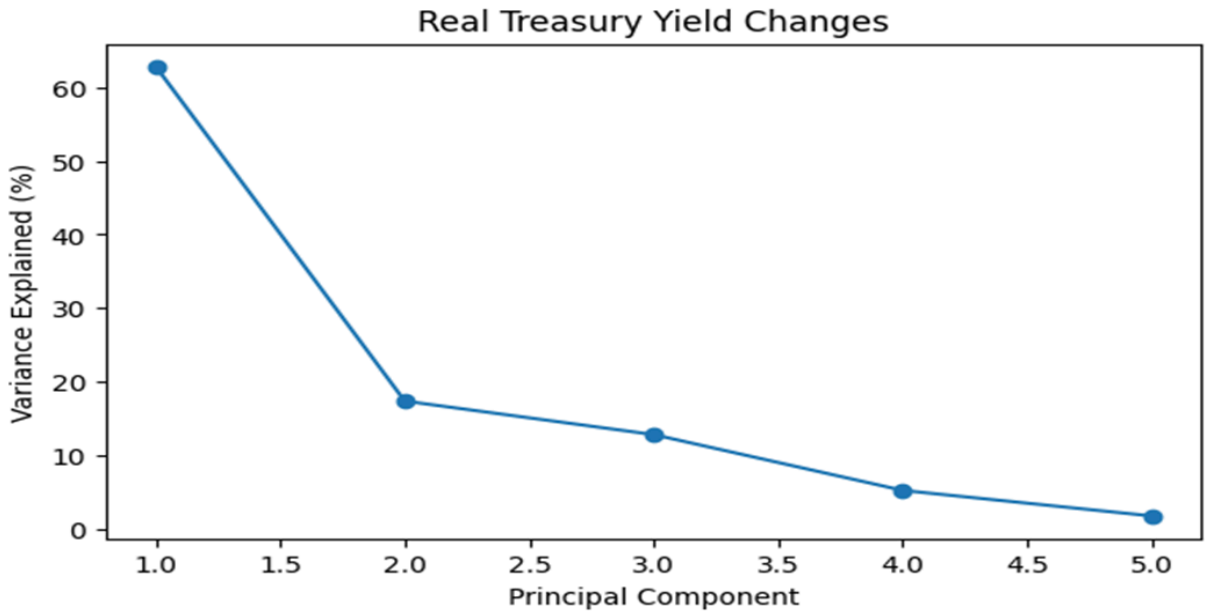
The analysis of PCA with respect to changes in the real Treasury yields shows that the first component explains the greatest proportion of the variance indicating the existence of high correlation amongst the

yields. Component 1 accounts for 62.84% of the variance (variance = 0.6284), Component 2 accounts for 17.38% (0.1738), Component 3 for 12.82% (0.1282), Component 4 for 5.20% (0.0520), and Component 5 accounts for 1.76% (0.0176).

The reason for the large variance explained by Component 1, is that it likely captures the dominant component of common movements (most likely parallel shifts in the yield curve) across the yields, with Component 2 reflecting possibly slope changes, and the subsequent components capturing predominantly noise. In contrast to simulation, where variance is distributed more evenly, the PCA clearly shows that the first component captures the majority of the variance, indicating lesser independence amongst the Treasury yields. In support of this, Tuckman and Serrat have suggested that the movements of Treasury yields are strongly influenced by fundamental economic changes.

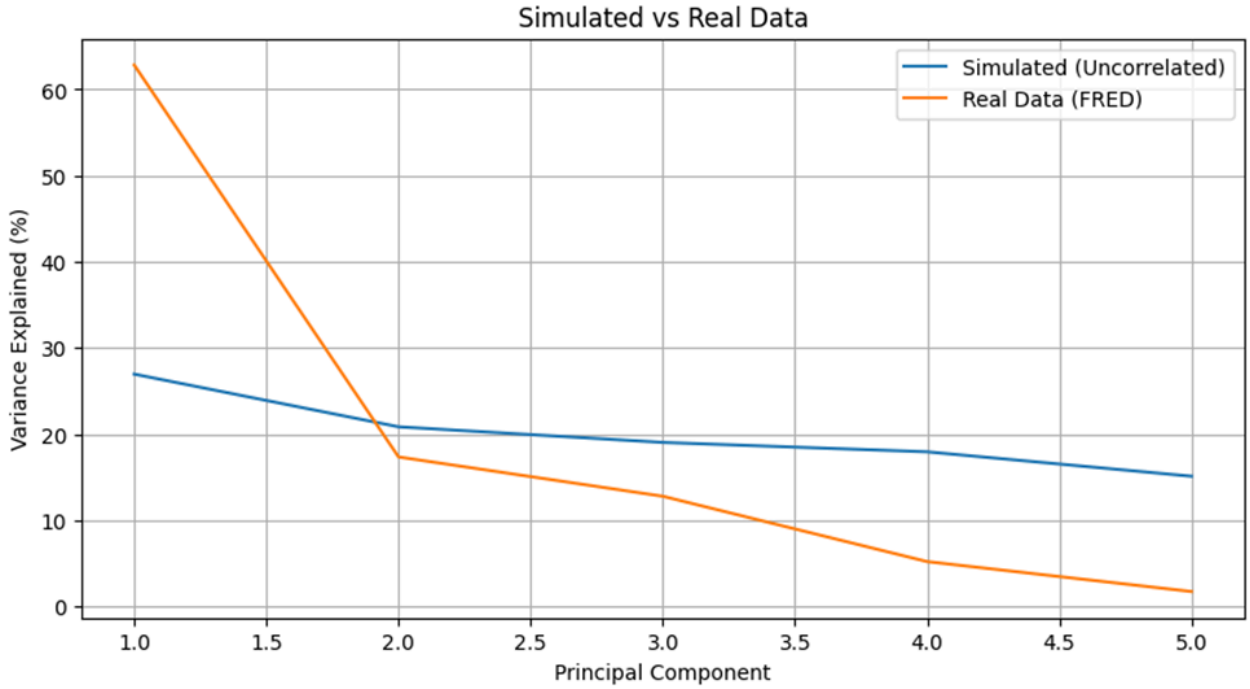
Component	Explained Variance	Variance Explained (%)
1	0.6284	62.84%
2	0.1738	17.38%
3	0.1282	12.82%
4	0.0520	5.20%
5	0.0176	1.76%

3i: Screeplot of Variance Explained for Each Component



The screeplot demonstrates a steep decline from Principal Component 1 to 5, where explained variance rations were **0.6284, 0.1738, 0.1282, 0.0520, and 0.0176.** The sharp drop from Component 1 to Component 2 indicates that **the first component captures the majority of the total variance**, while the later components contribute progressively smaller amounts. This steep decline reflects the high correlation among Treasury yields, in contrast to simulated data, where variance is more evenly distributed and no single component dominates.

3j: Comparing The Uncorrelated and Government Data Based Screeplots



The screeplot for the FCC unsupervised simulated datasets captures the explained variance ratios sequentially as 0.2698, 0.2087, 0.1906, 0.1797, and 0.1513 (26.98%, 20.87%, 19.06%, 17.97%, and 15.13%). The ratios confirm that the variances are fairly close and are similar in value.

The screeplot loses variance very fast after the first component for the actual government yields data. The explained variance for this dataset are 0.6284, 0.1738, 0.1282, 0.0520, and 0.0176 (62.84%, 17.38%, 12.82%, 5.20%, and 1.76%). The first component is attributed to a value of 62.84% dominance and the rest are relatively negligible (ChatGPT OpenAI, 2025).

The difference in variance capturing in the two scenarios is due to the unsupervised sample computed without correlation. The real sample of treasury yield data has high correlation and likely captures a common factor. This is most likely parallel shifts of the yield curve. The flat screeplot correlates to the unsupervised data and the filled screeplot demonstrates the difference in correlation as Jolliffe and Cadima discussed.

4. Empirical Analysis of ETFs

a,b,c-)

For this empirical analysis, I selected the Real Estate Select Sector SPDR Fund (ticker: XLRE). The primary motivation for this choice is its focus on the U.S. real estate sector, providing a pure-play exposure to Real Estate Investment Trusts (REITs) and other real estate companies. This sector is particularly interesting for analysis due to its unique risk-return profile, which is often driven by factors like interest rates, property market cycles, and rental income, distinct from the broader equity market. From a practical standpoint, XLRE's underlying holdings consist of large, liquid U.S. stocks, ensuring the availability of high-quality, reliable historical data. This data integrity is a fundamental requirement for the subsequent quantitative analysis, including portfolio construction and risk management techniques central to financial engineering (Tuckman and Serrat; James et al.).

The analysis begins with the constituents of the ETF. To ensure accuracy and avoid potential data parsing errors from automated web scraping, I manually compiled a list of the **30 largest holdings** of XLRE as of the analysis date. This list includes major industry leaders like **Prologis (PLD), American Tower (AMT), Equinix (EQIX), Welltower (WELL), Public Storage (PSA), and Simon Property Group (SPG)**.

The historical daily closing prices for these 30 assets were then programmatically downloaded using the "yfinance" library in Python (Ranatunga). The data period spans from **March 6, 2025, to October 6, 2025**, providing approximately six months of daily price data for the analysis. It's important to note that "yfinance" provides prices that are adjusted for corporate actions, meaning all data is already adjusted for dividends and stock splits, which is crucial for calculating accurate returns.

A quick snapshot of the raw data structure is shown below:

Date	AHH	AMT	ARE	AVB	BXP
2025-03-06	8.351662	201.859192	95.140060	212.911575	64.164467
2025-03-07	8.549837	207.087845	97.704453	212.862869	66.406906
2025-03-10	8.398846	207.848755	97.399376	210.836060	65.608101
2025-03-11	8.238419	204.639374	94.634796	207.493790	63.105808
2025-03-12	8.210108	201.625092	93.919815	204.667938	63.606258

The dataset contains around 120 trading days over six months.

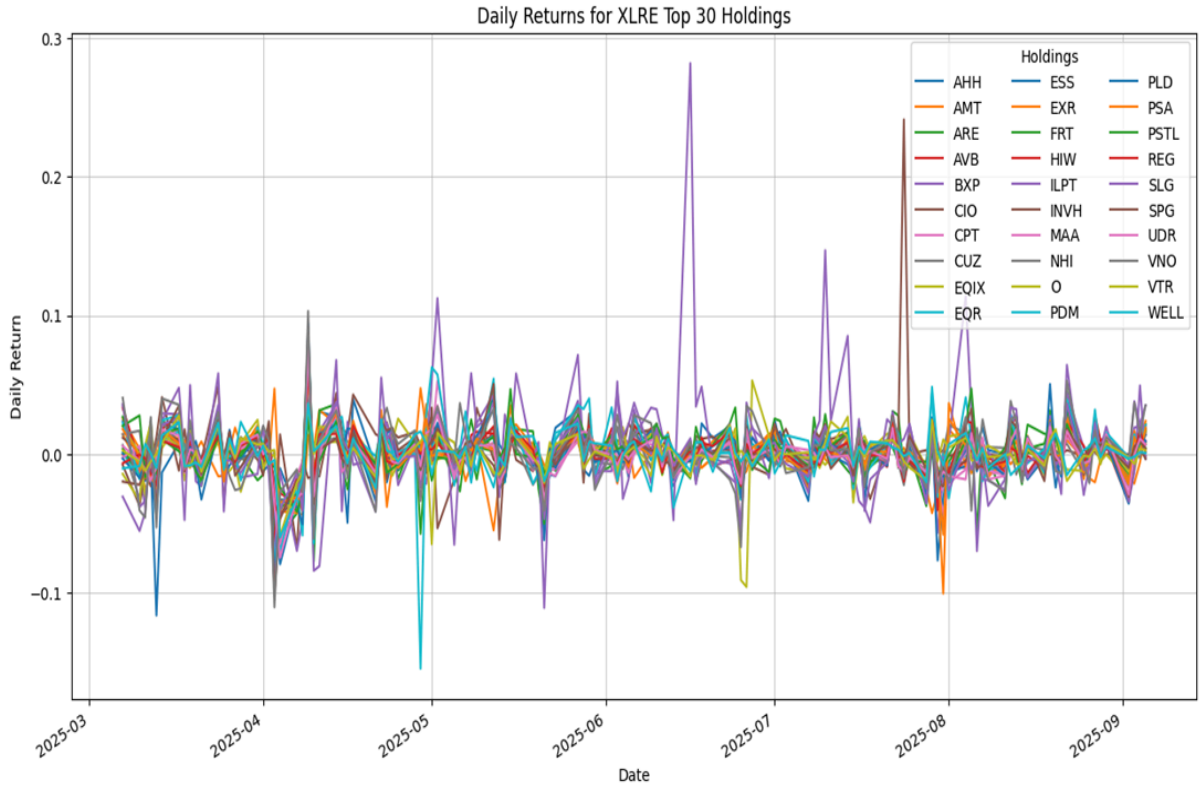
Daily returns were computed using the standard formula :

$$Daily\ Return\ = \frac{Price_t - Price_{t-1}}{Price_{t-1}}$$

The missing values are dropped to clean the dataset. A sample of the daily returns for the first few rows:

Date	AHH	AMT	ARE	AVB	BXP
2025-03-07	0.023729	0.025902	0.026954	-0.000229	0.034948
2025-03-10	-0.017660	0.003674	-0.003122	-0.009522	-0.012029
2025-03-11	-0.019101	-0.015441	-0.028384	-0.015852	-0.038140
2025-03-12	-0.003436	-0.014730	-0.007555	-0.013619	0.007930
2025-03-13	-0.116092	0.005225	-0.018981	-0.010379	-0.032077

The daily return time series captures both positive and negative fluctuations and allows for further statistical analysis. The volatility and correlations across companies are illustrated at a line plot of daily returns



The plot shows periods of both high and low volatility for 30 largest holdings, with some large swings corresponding to market events.

d.e.f)

Before we compute the covariance matrix, we attempt to standardize the data, by calculating **z-scores**.

$$Z\ Score_{Matrix} = \frac{Data\ Matrix - Mean\ Vector}{Standart\ Deviation\ Vector}$$

Covariance and correlation matrices were computed for the top holdings to quantify how returns move together:

Ticker	PLD	AMT	EQIX	WELL	PSA	SPG	O	AVB	EQR	INVH
PLD	1.000	0.101	0.606	0.425	0.691	0.821	0.484	0.735	0.756	0.697
AMT	0.101	1.000	0.165	0.346	0.441	-0.048	0.557	0.226	0.261	0.411
EQIX	0.606	0.165	1.000	0.395	0.500	0.555	0.333	0.534	0.525	0.473
WELL	0.425	0.346	0.395	1.000	0.574	0.372	0.590	0.523	0.545	0.567
PSA	0.691	0.441	0.500	0.574	1.000	0.565	0.727	0.788	0.769	0.785
SPG	0.821	-0.048	0.555	0.372	0.565	1.000	0.392	0.745	0.765	0.621
O	0.484	0.557	0.333	0.590	0.727	0.392	1.000	0.596	0.621	0.687
AVB	0.735	0.226	0.534	0.523	0.788	0.745	0.596	1.000	0.967	0.827
EQR	0.756	0.261	0.525	0.545	0.769	0.765	0.621	0.967	1.000	0.854
INVH	0.697	0.411	0.473	0.567	0.785	0.621	0.687	0.827	0.854	1.000

The correlation matrix shows how the daily returns of the top 10 holdings in the XLRE (Real Estate Select Sector SPDR Fund) move together. Overall, the correlations are mostly positive and fairly strong, which is expected since these companies all belong to the real estate sector. The highest correlation appears between **AVB and EQR (0.967)**, both residential REITs, which makes sense given their similar exposure to the multifamily housing market. **PLD and SPG (0.821)**, as well as **PSA and O (0.727)**,

also show strong relationships, likely because they are influenced by similar factors like interest rate movements and demand in the property market.

On the other hand, **AMT and SPG (-0.048)** or **AMT and EQIX (0.165)** have much weaker correlations, suggesting that they operate in different segments which adds some diversification within the portfolio.

In short, while the holdings tend to move in the same direction due to sector-wide effects, there is still some diversification, especially between specialized and residential REITs.

In quantitative finance, it's crucial to understand what drives the movements of asset returns. This understanding forms the foundation of sound portfolio management and risk optimization. Two widely used techniques that help uncover these underlying structures are **Principal Component Analysis (PCA)** and **Singular Value Decomposition (SVD)**.

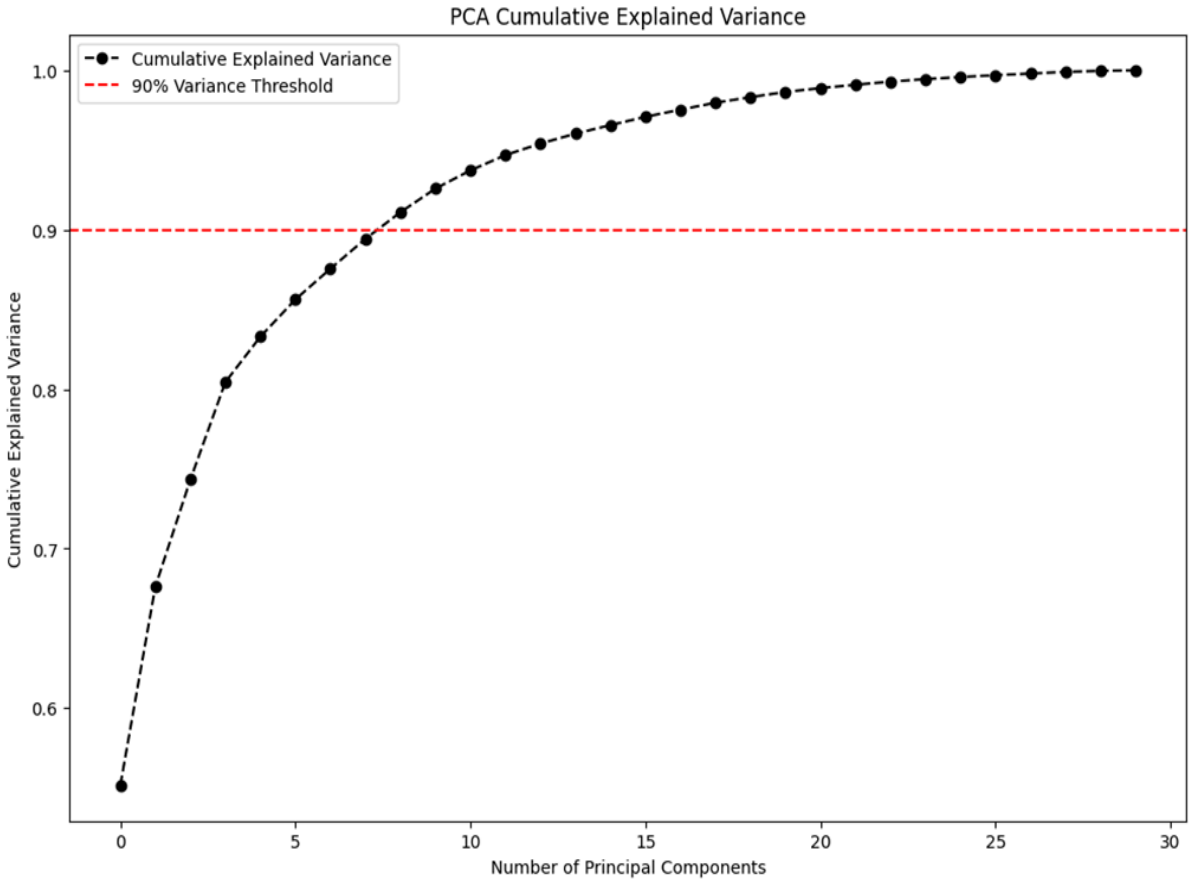
Both methods simplify complex financial data by reducing its dimensionality, while still keeping most of the important information intact. In simple terms, **PCA** breaks down the covariance matrix of asset returns

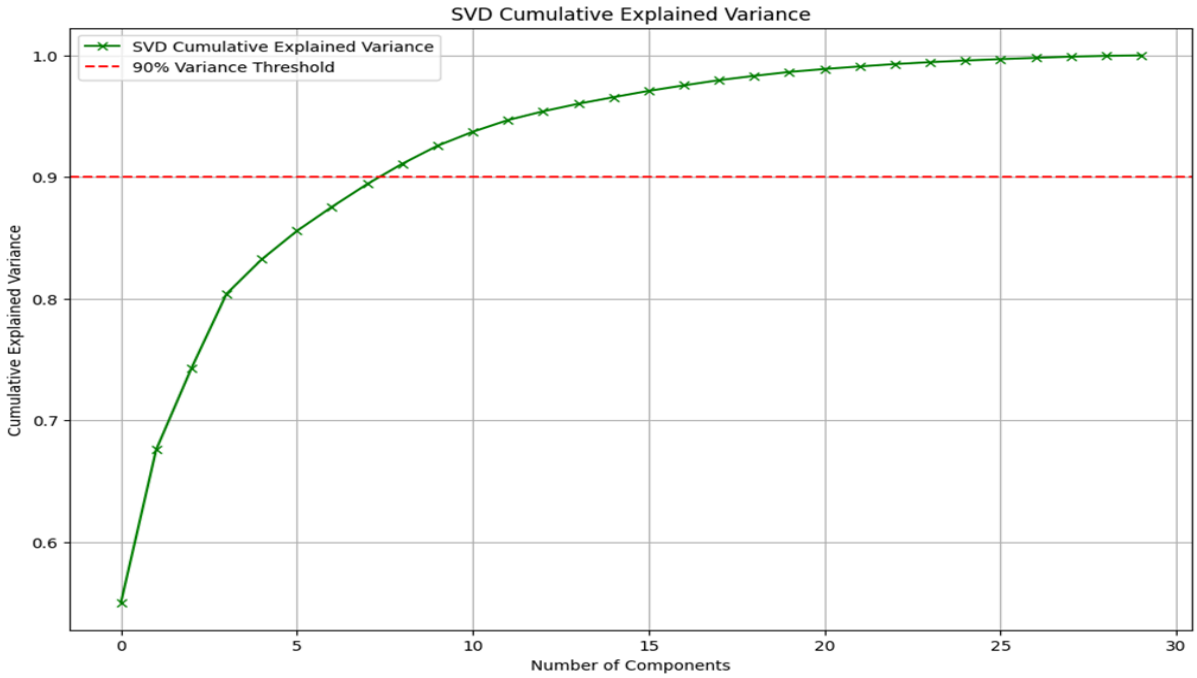
into a set of uncorrelated components, known as *principal components*. Each of these components represents a distinct source of variation in the data with the first component explaining the largest share of overall variance, and each following one contributing slightly less.

Component	PCA (Explained Variance)	SVD (Explained Variance)
PC1	0.5516	0.5501
PC2	0.1246	0.1256
PC3	0.0670	0.0668
PC4	0.0612	0.0614
PC5	0.0285	0.0284
PC6	0.0234	0.0233
PC7	0.0193	0.0199
PC8	0.0187	0.0187
PC9	0.0167	0.0167
PC10	0.0148	0.0148

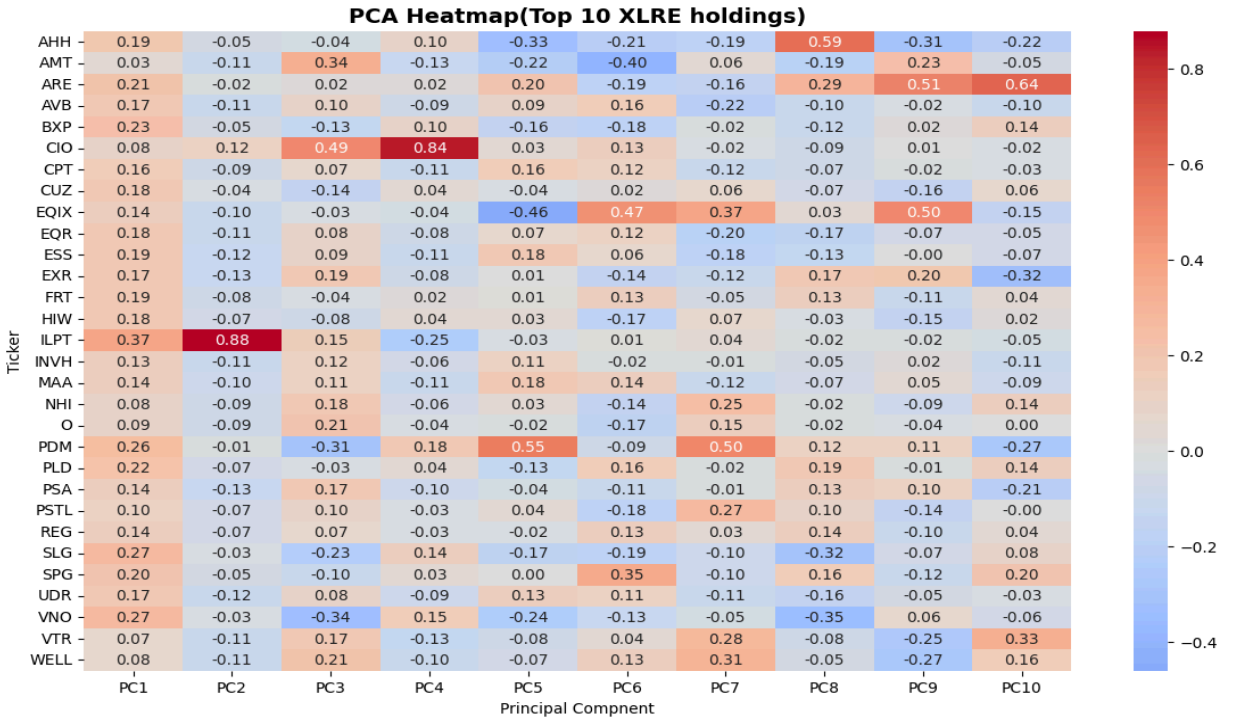
The results confirm a **perfect agreement** between PCA and SVD methods, as the explained variance values are nearly identical for each of the first ten components. The **first principal component (PC1)** alone explains about **55% of the total variance**, indicating the presence of a strong **common market factor** influencing all real estate stocks within the XLRE portfolio. After that PC2 explains 12% of the total variance.

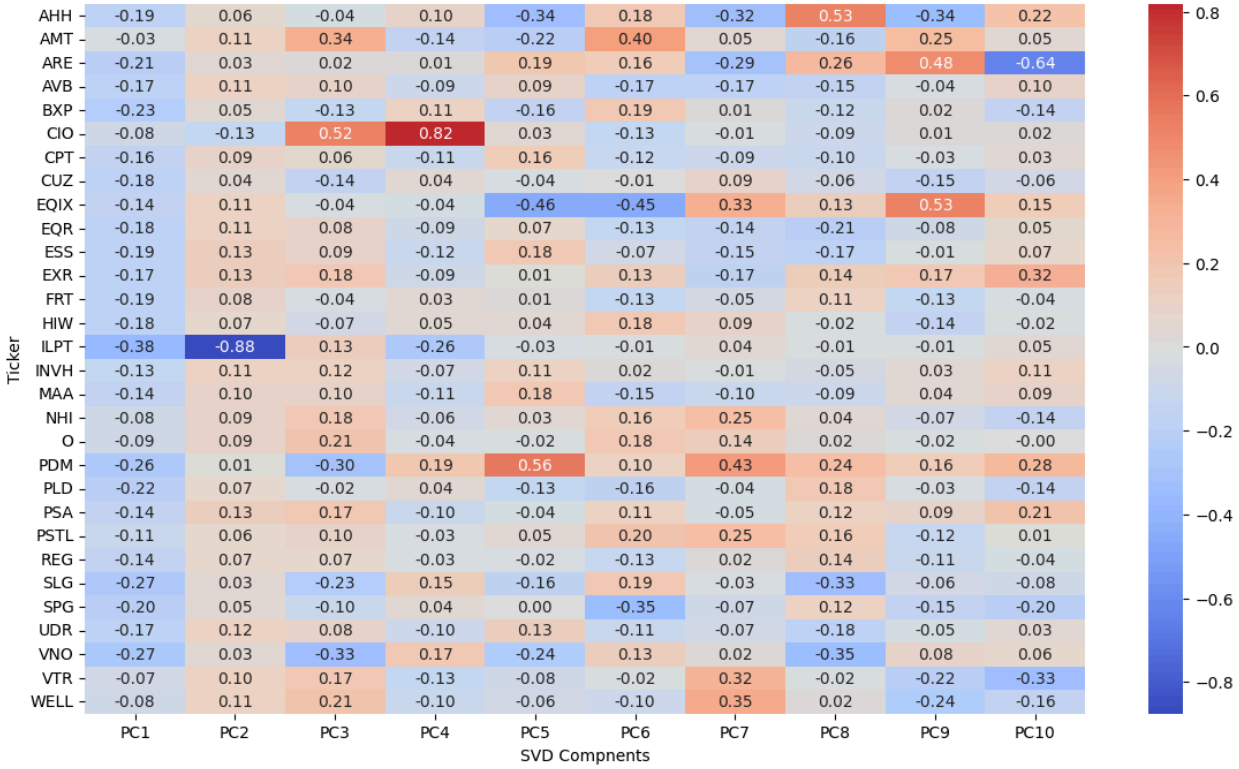
From the **cumulative variance curve (PCA and SVD)** below, we can see that **around eight principal components** are sufficient to capture roughly **91% of the total variance** in the dataset. This highlights that, although the holdings are within the same sector, there is still a certain degree of intra-sector heterogeneity for example, between specialized REITs (such as data centers or communication infrastructure) and traditional residential or commercial REITs.





Next, we present heatmaps illustrating the individual loadings of each stock on the principal components. At first glance, some variations in color may suggest differences between PCA and SVD outputs, however, a closer inspection reveals that these are simply due to sign flips. This phenomenon does not affect the interpretation, as principal components represent directions in space rather than absolute values. A change in sign merely reverses the vector along the same line, leaving the underlying relationships intact. Consequently, the heatmaps confirm that both analyses are consistent and mathematically equivalent (Jolliffe and Cadima).


PCA Heatmap(Top 10 XLRE holdings)
Ticker
AHH 0.19 -0.05 -0.04 0.10 -0.33 -0.21 -0.19 0.59 -0.31 -0.22
AMT 0.03 -0.11 0.34 -0.13 -0.22 -0.40 0.06 -0.19 0.23 -0.05
ARE 0.21 -0.02 0.02 0.02 0.20 -0.19 -0.16 0.29 0.51 0.64
AVB 0.17 -0.11 0.10 -0.09 0.09 0.16 -0.22 -0.10 -0.02 -0.10
BXP 0.23 -0.05 -0.13 0.10 -0.16 -0.18 -0.02 -0.12 0.02 0.14
CIO 0.08 0.12 0.49 0.84 0.03 0.13 -0.02 -0.09 0.01 -0.02
CPT 0.16 -0.09 0.07 -0.11 0.16 0.12 -0.12 -0.07 -0.02 -0.03
CUZ 0.18 -0.04 -0.14 0.04 -0.04 0.02 0.06 -0.07 -0.16 0.06
EQIX 0.14 -0.10 -0.03 -0.04 -0.46 0.47 0.37 0.03 0.50 -0.15
EQR 0.18 -0.11 0.08 -0.08 0.07 0.12 -0.20 -0.17 -0.07 -0.05
ESS 0.19 -0.12 0.09 -0.11 0.18 0.06 -0.18 -0.13 -0.00 -0.07
EXR 0.17 -0.13 0.19 -0.08 0.01 -0.14 -0.12 0.17 0.20 -0.32
FRT 0.19 -0.08 -0.04 0.02 0.01 0.13 -0.05 0.13 -0.11 0.04
HIW 0.18 -0.07 -0.08 0.04 0.03 -0.17 0.07 -0.03 -0.15 0.02
ILPT 0.37 0.88 0.15 -0.25 -0.03 0.01 0.04 -0.02 -0.02 -0.05
INVH 0.13 -0.11 0.12 -0.06 0.11 -0.02 -0.01 -0.05 0.02 -0.11
MAA 0.14 -0.10 0.11 -0.11 0.18 0.14 -0.12 -0.07 0.05 -0.09
NHI 0.08 -0.09 0.18 -0.06 0.03 -0.14 0.25 -0.02 -0.09 0.14
O 0.09 -0.09 0.21 -0.04 -0.02 -0.17 0.15 -0.02 -0.04 0.00
PDM 0.26 -0.01 -0.31 0.18 0.55 -0.09 0.50 0.12 0.11 -0.27
PLD 0.22 -0.07 -0.03 0.04 -0.13 0.16 -0.02 0.19 -0.01 0.14
PSA 0.14 -0.13 0.17 -0.10 -0.04 -0.11 -0.01 0.13 0.10 -0.21
PSTL 0.10 -0.07 0.10 -0.03 0.04 -0.18 0.27 0.10 -0.14 -0.00
REG 0.14 -0.07 0.07 -0.03 -0.02 0.13 0.03 0.14 -0.10 0.04
SLG 0.27 -0.03 -0.23 0.14 -0.17 -0.19 -0.10 -0.32 -0.07 0.08
SPG 0.20 -0.05 -0.10 0.03 0.00 0.35 -0.10 0.16 -0.12 0.20
UDR 0.17 -0.12 0.08 -0.09 0.13 0.11 -0.11 -0.16 -0.05 -0.03
VNO 0.27 -0.03 -0.34 0.15 -0.24 -0.13 -0.05 -0.35 0.06 -0.06
VTR 0.07 -0.11 0.17 -0.13 -0.08 0.04 0.28 -0.08 -0.25 0.33
WELL 0.08 -0.11 0.21 -0.10 -0.07 0.13 0.31 -0.05 -0.27 0.16
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10
Principal Compnent
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


SVD Heatmap(Top 10 XLRE holding
Ticker
AHH -0.19 0.06 -0.04 0.10 -0.34 0.18 -0.32 0.53 -0.34 0.22
AMT -0.03 0.11 0.34 -0.14 -0.22 0.40 0.05 -0.16 0.25 0.05
ARE -0.21 0.03 0.02 0.01 0.19 0.16 -0.29 0.26 0.48 -0.64
AVB -0.17 0.11 0.10 -0.09 0.09 -0.17 -0.17 -0.15 -0.04 0.10
BXP -0.23 0.05 -0.13 0.11 -0.16 0.19 0.01 -0.12 0.02 -0.14
CIO -0.08 -0.13 0.52 0.82 0.03 -0.13 -0.01 -0.09 0.01 0.02
CPT -0.16 0.09 0.06 -0.11 0.16 -0.12 -0.09 -0.10 -0.03 0.03
CUZ -0.18 0.04 -0.14 0.04 -0.04 -0.01 0.09 -0.06 -0.15 -0.06
EQIX -0.14 0.11 -0.04 -0.04 -0.46 -0.45 0.33 0.13 0.53 0.15
EQR -0.18 0.11 0.08 -0.09 0.07 -0.13 -0.14 -0.21 -0.08 0.05
ESS -0.19 0.13 0.09 -0.12 0.18 -0.07 -0.15 -0.17 -0.01 0.07
EXR -0.17 0.13 0.18 -0.09 0.01 0.13 -0.17 0.14 0.17 0.32
FRT -0.19 0.08 -0.04 0.03 0.01 -0.13 -0.05 0.11 -0.13 -0.04
HIW -0.18 0.07 -0.07 0.05 0.04 0.18 0.09 -0.02 -0.14 -0.02
ILPT -0.38 -0.88 0.13 -0.26 -0.03 -0.01 0.04 -0.01 -0.01 0.05
INVH -0.13 0.11 0.12 -0.07 0.11 0.02 -0.01 -0.05 0.03 0.11
MAA -0.14 0.10 0.10 -0.11 0.18 -0.15 -0.10 -0.09 0.04 0.09
NHI -0.08 0.09 0.18 -0.06 0.03 0.16 0.25 0.04 -0.07 -0.14
O -0.09 0.09 0.21 -0.04 -0.02 0.18 0.14 0.02 -0.02 -0.00
PDM -0.26 0.01 -0.30 0.19 0.56 0.10 0.43 0.24 0.16 0.28
PLD -0.22 0.07 -0.02 0.04 -0.13 -0.16 -0.04 0.18 -0.03 -0.14
PSA -0.14 0.13 0.17 -0.10 -0.04 0.11 -0.05 0.12 0.09 0.21
PSTL -0.11 0.06 0.10 -0.03 0.05 0.20 0.25 0.16 -0.12 0.01
REG -0.14 0.07 0.07 -0.03 -0.02 -0.13 0.02 0.14 -0.11 -0.04
SLG -0.27 0.03 -0.23 0.15 -0.16 0.19 -0.03 -0.33 -0.06 -0.08
SPG -0.20 0.05 -0.10 0.04 0.00 -0.35 -0.07 0.12 -0.15 -0.20
UDR -0.17 0.12 0.08 -0.10 0.13 -0.11 -0.07 -0.18 -0.05 0.03
VNO -0.27 0.03 -0.33 0.17 -0.24 0.13 0.02 -0.35 0.08 0.06
VTR -0.07 0.10 0.17 -0.13 -0.08 -0.02 0.32 -0.02 -0.22 -0.33
WELL -0.08 0.11 0.21 -0.10 -0.06 -0.10 0.35 0.02 -0.24 -0.16
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10
SVD Compnents
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8

Reference

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer, 2006.

Federal Reserve Bank of St. Louis. "Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity, Quoted on an Investment Basis (DGS2)." *FRED*, https://fred.stlouisfed.org/series/DGS2. Accessed 08 October. 2025.

Bodie, Zvi, Alex Kane, and Alan J. Marcus. *Investments*. 12th ed., McGraw-Hill Education, 2021.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer, 2009.

Jolliffe, Ian T., and Jürgen Cadima. "Principal Component Analysis: A Review and Recent Developments." *Philosophical Transactions of the Society: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016, pp. 20150202. Society, doi:10.1098/rsta.2015.0202

State Street Global Advisors. (s.d.). *Real Estate Select Sector SPDR Fund (XLRE) holdings*. https://www.ssga.com/us/en/intermediary/etfs/funds/real-estate-select-sector-spdr-fund-xlre