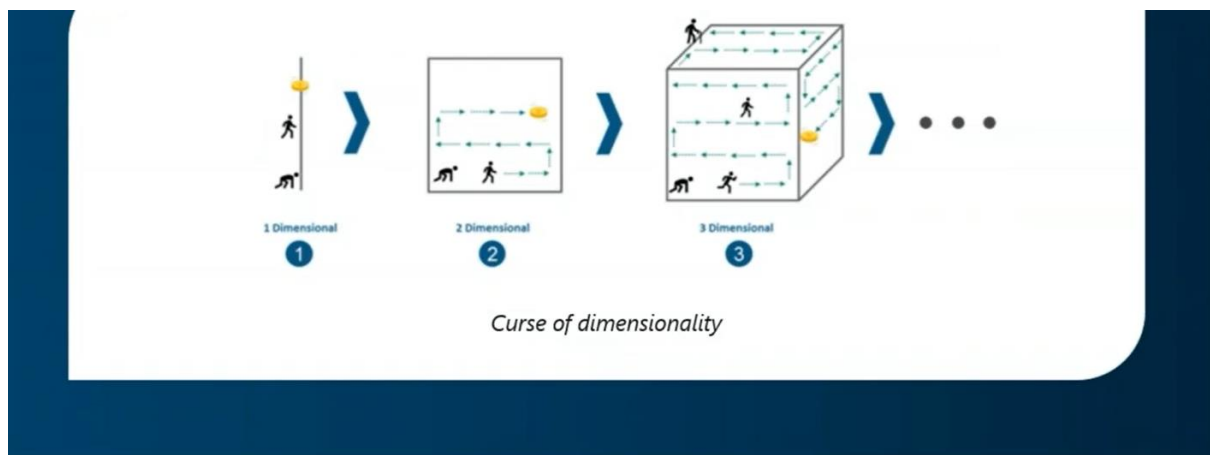


# Need for Principle Component Analysis

## What is PCA?

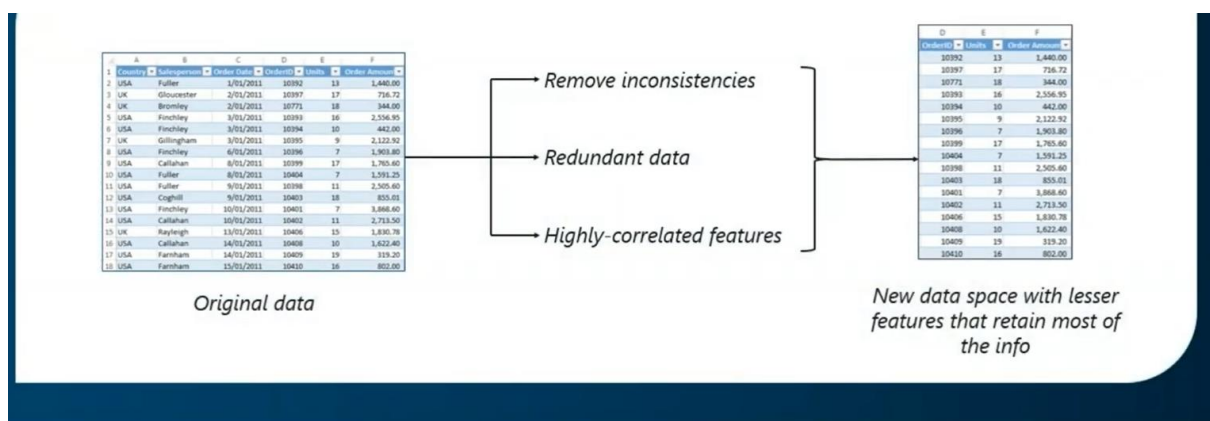
### Step by Step Computation of PCA

### Principle Component Analysis with Python



### Need For PCA

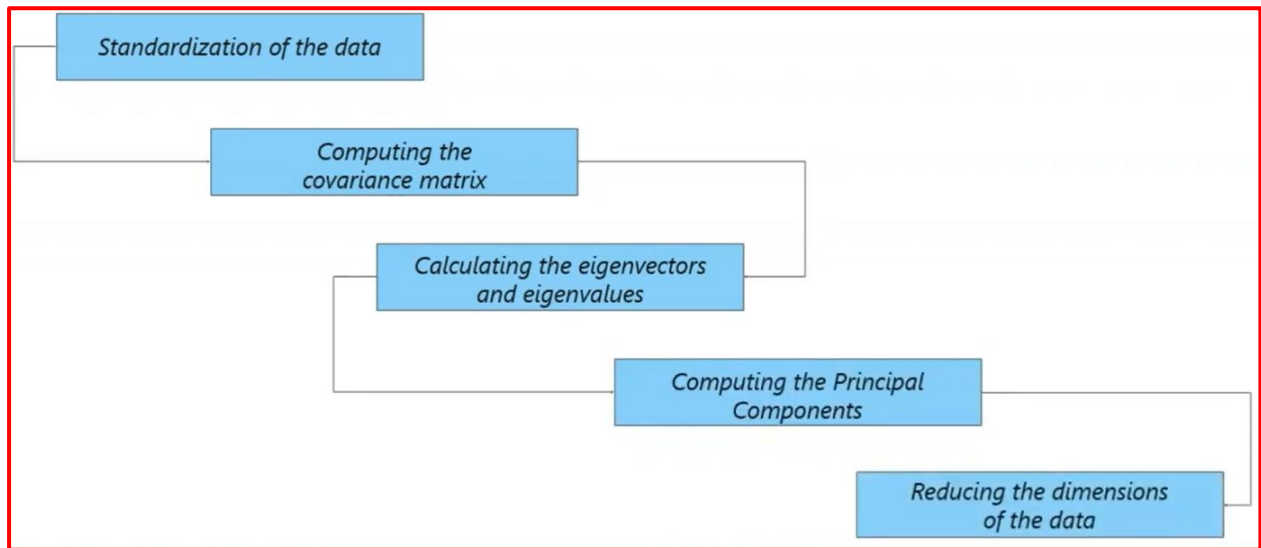
High dimension data is extremely complex to process due to inconsistencies in the features which increase the computation time and make data processing and EDA more convoluted.



### What Is Principal Component Analysis (PCA)?

Principal components analysis (PCA) is a dimensionality reduction technique that enables you to identify correlations and patterns in a data set so that it can be transformed into a data set of significantly lower dimension without loss of any important information.

# STEP BY STEP PCA



Rating	# of downloads
5	1383
3	668
2	763
5	839
1	342

Rating feature ranges between 0-5

# of downloads ranges between 100-5000

$$Z = \frac{\text{Variable value} - \text{mean}}{\text{Standard deviation}}$$

## Step 1: Standardization of the data

Standardization is all about scaling your data in such a way that all the variables and their values lie within a similar range.

$$\begin{bmatrix} \text{Cov}(a, a) & \text{Cov}(a, b) \\ \text{Cov}(b, a) & \text{Cov}(b, b) \end{bmatrix}$$

The covariance value denotes:

- How co-dependent two variables are
- Negative covariance - indirectly proportional
- positive covariance - directly proportional to

### Step 2: Computing the covariance matrix

*A covariance matrix expresses the correlation between the different variables in the data set. It is essential to identify heavily dependent variables because they contain biased and redundant information which reduces the overall performance of the model.*

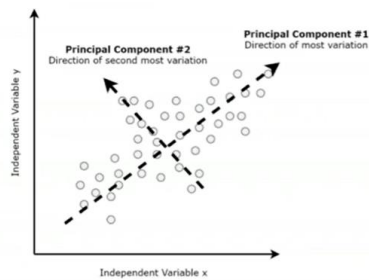
*Principal components are the new set of variables that are obtained from the initial set of variables. They compress and possess most of the useful information that was scattered among the initial variables.*

- *Eigenvectors are those vectors when a linear transformation is performed on them then their direction does not change.*
- *Eigenvalues simply denote the scalars of the respective eigenvectors*

### Step 3: Calculating the Eigenvectors and Eigenvalues

*Eigenvectors and eigenvalues are the mathematical constructs that must be computed from the covariance matrix in order to determine the principal components of the data set.*

- *PC1 is the most significant and stores the maximum possible information.*
- *PC2 is the second most significant PC and stores remaining maximum info and so on*



#### **Step 4: Computing the Principal Components**

*Once we have computed the Eigenvectors and eigenvalues, all we have to do is order them in the descending order, where the eigenvector with the highest eigenvalue is the most significant and thus forms the first principal component.*