1. The general idea of principal component analysis (PCA) and its usefulness in data mining systems.

For datasets with many variables, there will be too many pairwise correlations between the variables to consider. PCA reduces the number of variables to a few, interpretable linear combinations of the data. Each linear combination will correspond to a principal component.

In data mining systems, the accuracy and reliability of a classification or prediction model will suffer if we include highly correlated variables or variables that are unrelated to the outcome of interest because of over-fitting. PCA helps reduce dimensionality without sacrificing accuracy and allows the user to spot trends, patterns and outliers in the data in an easier way.

1. See lab2\_PCA.m section II-Derive principal components
2. See lab2\_PCA.m section III-The use of pca function

Coeff contains the coefficients for the principal components;

Score contains the actual principal components;

Latent stores the variance of the respective principal components.

Score = normalized dataset \* rearranged eigenvectors of the covariance matrix

1. The first column accounts for as much as the variability, and the others account for the rest.
2. See lab2\_PCA.m section IV- Plot the cumulative contribution of the PCs to the variance
3. The first principal component contains nearly 98% of the variance of the original table. So I would compress the 11 variables into 1, while losing only 2% of the variance.