IKEA:

1. Clean the dataset - check the data types, exclude the categorical data/textual.
2. Perform PCA for dimensionality reduction (probably explain 90-95% of the total variance) – normalise/scale the data.
3. Apply the Elbow method to select an optimum number of clusters – it would turn out to be 3.
4. Apply k-means (feed the results from PCA to the induction model).
5. Append cluster labels column to the original df.
6. Append a column “has_ikea” to the original df.
7. Explore the clusters - you will probably find them separated into small, medium, and large municipalities.
8. Analyse them based on, for example, population, sales index, revenue etc., exclude where IKEA already exists or is close to existing IKEA.
9. Suggest top 3 candidates for opening new IKEA stores.

Titanic:

(Hint: The most significant task here is preparing the data.)

1. Explore the dataset (understand the features and make some basic assumptions).
2. Think about what features would affect the survival rate (you would probably want to drop the ones that do not). It may be risky, though, so you could opt for performing a Correlation analysis to see which features are most correlated to survival. Then you are not relying so much on guessing.
3. Pre-process: Handling missing data, converting features (to numeric), creating categories, creating new features.
4. Fit the classifier/evaluate performance using the k-fold cross-validation.
5. Analyse.