**Step 1 - Create a new project**

**Step 2 - Import Datasets to dataiku workflow**

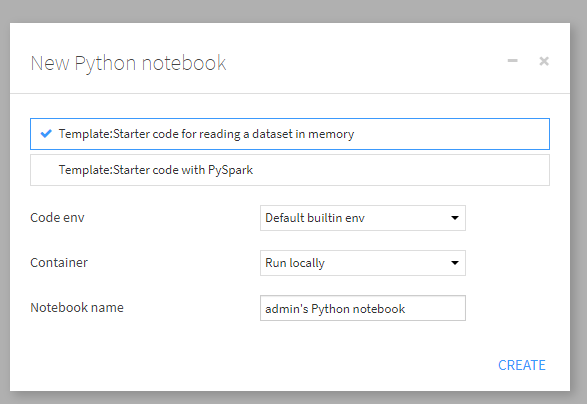
1. Import “text8.csv” document via “Datasets” menu
   1. New Datasets → Upload Files → Drag & Drop “data/text8.csv”
2. Add “text8” DSS dataset to workflow
3. Navigate to Workflow → create a visual recipe to clean the dataset “data preparation”
   1. Natural Language → Simplify Text

**Step 3 - Data Understanding/Exploration**

1. Explore data via “Visual Analysis” and/or “Lab”

**Step 4 - Model Building Process (CRISP-DM) & Scikit-Learn Library**

1. Create new jupyter notebook in Dataiku
   1. Notebooks → Create Notebook → Python



1. Import Dataset & Re-Create previously created python code
   1. Workbook: Scikit-Learn - Dataiku LDA Model (7).ipynb
2. execute notebook - ensure components and model build works correctly

**Step 5 - Additional Model Build**

1. Create a Python Code Recipe that produces LDA Model - goto “Lab” on workflow
2. Choose “Clustering” task, since LDA creates unlabeled themes for classifying text
3. Choose “Quick Models” - “Interactive Clustering’ - Train
4. Need to use text for clustering - goto “Models” - Design - Features handling - select “text1”
5. Use Term Hashing + SVD - select “Train” and evaluate results
6. Create name for model for use in workflow, then “deploy model”
7. Follow the steps in the “deploy model” section, to add it to the workflow

**Class 6 - Model Comparison & Evaluation**

1. CLick - “Interactive Clustering” to evaluate model results
   1. Includes: Summary, Var importance, Clusters & Model Informaiton
2. Build Multiple Models in “visual analysis”, then select the best model
3. Compare “best” model vs. LDA Model from juypter notebook - is it better/worse?

**Class 7 - Model Deployment & Operalization**

1. Publish the model results to a dashboard, to monitor future results
2. Schedule and Distribute the resulting analysis dataset, for use in data products (such as tableau)
   1. In this case, download the “scored” dataset to CSV, for use in other dashboards.