Data Science Workflow

1. **Business Problem Understanding**:
   - Align data science objectives with business goals to ensure the solution addresses the right problem.
2. **Data Collection**:
   - Gather relevant data from varied sources to get a comprehensive dataset for accurate analysis.
3. **Data Cleaning**:
   - **Treat Null Values**
   - **Filter Erroneous Outliers**
   - **Eliminate Duplicates**
4. **Exploratory Data Analysis (EDA)**:
   - Conduct thorough statistical and visual analysis to understand the data, uncover patterns, trends, and anomalies in the data.
5. **Feature Engineering**:
   - **Prune Redundant Features**
   - **Encode Categorical Features**
   - **Invent New Features**
6. **Optimize Machine Learning (ML) Process**:
   - **Train/Test Split**
   - **Cross-validation**: Use cross-validation on the train set to validate various models and hyperparameters.
   - **Model Selection**: Use the validation performance to choose a few good models, then confirm their performance on the test set.
   - **Validate Model Robustness**: Select the best model from the previous step. Perform extensive cross-validation with a big number of splits and with all the data in order to assess production readiness.
   - **Finalize Model Training**: If the previous step is successful, train the final model on the entire dataset.
   - **Efficient Model Storage**: Save the model for future use.
7. **Deploy Model in Production**:
   - Deploy the stored model in a real-world environment, monitor its performance, and maintain it for accuracy and reliability.