from graphviz import Digraph

# Create a new Digraph

dot = Digraph(comment='Text Data Analysis Workflow')

# Add nodes

dot.node('A', 'Data Collection\nGather text data from sources like social media, product reviews, news articles, and customer feedback forms during the product launch.')

dot.node('B', 'Data Cleaning\nPreprocess the collected text: remove noise, stopwords, duplicates, special characters, and handle missing data for consistency.')

dot.node('C', 'Exploratory Data Analysis (EDA)\nDiscover patterns in text data such as frequency of terms, initial sentiment direction, and activity over time.')

dot.node('D', 'Feature Extraction\nUse NLP techniques to extract features such as sentiment scores, key phrases, named entities, and product-related keywords.')

dot.node('E', 'Data Visualization\nPlot sentiment over time, display word clouds, bar charts of frequently mentioned terms, and other visualizations.')

dot.node('F', 'Reporting\nCompile insights into a structured format like a dashboard or presentation for product teams and decision-makers.')

# Connect the nodes

dot.edges(['AB', 'BC', 'CD', 'DE', 'EF'])

# Render and display

dot.render('text\_data\_analysis\_workflow', format='png', view=True)