from graphviz import Digraph

dot = Digraph(format='png')

dot.attr(rankdir='TB', bgcolor='white', style='filled', fillcolor='#E6F4F1')

steps = {

"Data Collection": "Use Twitter API or scraping tools to collect tweets using relevant hashtags, mentions, and keywords around the product launch.",

"Data Cleaning": "Remove retweets, spam, URLs, emojis, and irrelevant tweets. Normalize text (lowercase, remove punctuation, etc.).",

"Exploratory Data Analysis (EDA)": "Analyze tweet volume over time, common words, user engagement metrics (likes, retweets), and initial sentiment distribution.",

"Feature Extraction": "Apply NLP models to extract sentiment (positive, negative, neutral), detect trending hashtags, topics, user locations, and influential users.",

"Data Visualization": "Generate sentiment timelines, word clouds, bar charts of top hashtags, geolocation maps, and engagement heatmaps.",

"Reporting": "Summarize insights for stakeholders using dashboards, reports, and presentations to inform marketing or product strategy."

}

# Add nodes

for i, (title, description) in enumerate(steps.items()):

node\_id = f"step{i}"

label = f"<<b>{title}</b><br/>{description}>>"

dot.node(node\_id, label=label, shape='box', style='filled', fillcolor='lightblue', fontsize='10', fontname='Helvetica', margin='0.3')

# Add edges

for i in range(len(steps) - 1):

dot.edge(f"step{i}", f"step{i+1}")

# Render the diagram

dot.render('twitter\_data\_analysis\_flowchart', view=True)