Data analysis pipeline