1. **Business Problem Identification:** Your ability to take a public dataset, read about it and then formulate a few hypotheses or a few made up business problems is going to demonstrate your ability to apply your knowledge to real-world applications. For example, in [this public e-commerce dataset](https://app.noteable.io/f/9d877fbd-4bb4-41ca-bf11-c7f48f0ccc20/Exploratory-Data-Analysis-Using-SQL-and-Python---Online-Retailer-Orders.ipynb), you could come up with a business problem of “Find the highest selling products and determine whether to advertise them on social media or not”
2. **Data Extraction**Querying, Cleaning, Aggregating messy datasets using SQL, Pandas (and whatever other libraries are your favorite). I can’t emphasize enough about how important it is to **showcase SQL skills**. As a hiring manager, it is highly unlikely that I will hire a data scientist who is great in Python but doesn’t know SQL.  
   - **Write SQL on Excel and CSV files directly**: Writing SQL in Jupyter Notebooks is a pain. You typically have to use a library and then write the SQL in quotes. **I recommend using a modern notebook like**[**noteable.io**](https://www.noteable.io/?utm_campaign=article-DS-portfolio-20221116) that allows you **run queries directly on an Excel or CSV file without needing any database**. After all, since are not applying for a Database Architect role, there really is little to no value in struggling to create a local database.  
   - **Connect to Database:**Of course, you can connect to databases in Noteable but it may not be that easy (or free) for you to get to access a public database. If you do have access (Google: “BigQuery public datasets”) to a database, then you will be able to write native SQL in Noteable. No need for packages, quotes and all that nonsense.  
   - **Create a Local DB:** Alternatively, you could create a local database on your computer but that all takes time and forces you to come up with data that you need to enter into the tables. Afterall, your goal is to showcase your SQL skills, not SQL Database Administration skills. **So, I consider this route a fairly big waste of time**.
3. **Data Exploration, Visualization**It is essential for any data practitioner to be able to write good quality code to quickly segment, filter, explore and visualize data. Most will use Python or R. So, while modern notebook platforms like Noteable will let you create charts interactively, I still recommend throwing in a couple of really cool visualizations (see r/dataisbeautiful for inspiration) that showcase your ability to use Pandas, lambda functions (optional but great), and matplotlib/seaborn etc.
4. **Machine Learning**However, most data scientists, data analysts, data engineers, business intelligence engineers and machine learning engineers will find it beneficial to include some of the following methods in their projects:  
   ● Decision Trees  
   ● K-means clustering  
   ● K-nearest neighbors (KNN)  
   ● Linear Regression  
   ● Logistic Regression  
   ● Naive Bayes  
   ● Principal Component Analysis (PCA)  
   ● Random Forests  
   ● Support Vector Machine  
   This list includes some of the most popular and most widely used methods in the professional world, and is therefore a great source of inspiration for your future projects. By having intimate knowledge with even just a few of these methods, you increase your chances significantly for impressing the hiring manager at your next job interview. I recommend that you consider the type of job you would like to get, and then work with some of the algorithms above that are commonly implemented in your chosen field.