**5 Easy Projects (Beginner)**

These focus on fundamental skills: data wrangling, visualization, and basic machine learning.

1. **Data Cleaning & EDA (Exploratory Data Analysis) on a Real Dataset**
   * Use Pandas, NumPy, Matplotlib, and Seaborn.
   * Choose datasets like Titanic, Netflix Movies, or Airbnb listings.
   * Focus on handling missing values, outliers, and basic insights.
2. **SQL Querying on a Public Database**
   * Use SQLite/MySQL/PostgreSQL.
   * Work with datasets like IMDb, World Bank, or NBA stats.
   * Perform data extraction, filtering, aggregation, and joins.
3. **Basic Web Scraping with BeautifulSoup & Selenium**
   * Scrape job postings, news articles, or Amazon product data.
   * Store results in a CSV or SQL database.
4. **Simple Linear Regression Model**
   * Predict house prices, sales, or stock trends.
   * Use Scikit-learn to train/test a model.
   * Explain key concepts like RMSE, MAE, and R-squared.
5. **Git & GitHub Practice with a Mini Data Science Report**
   * Create a GitHub repository, add a Jupyter Notebook, and document findings.
   * Use version control, branches, and pull requests.

**5 Intermediate Projects**

These reinforce SQL, feature engineering, and more advanced modeling.

1. **Data Pipelines with Pandas & SQL**
   * Automate data extraction from CSV/API, clean it, and load it into an SQL database.
   * Use Pandas and SQLAlchemy.
2. **Time Series Forecasting (Stock Prices or Sales)**
   * Use ARIMA, Exponential Smoothing, or Facebook Prophet.
   * Work with daily/weekly stock prices, COVID-19 cases, or store sales.
3. **End-to-End Machine Learning Project (Classification or Regression)**
   * Use a dataset like Titanic, Loan Prediction, or Customer Churn.
   * Perform feature engineering, hyperparameter tuning, and model evaluation.
4. **Building a Flask API for a Machine Learning Model**
   * Train a model and deploy it using Flask.
   * Create an API endpoint to accept input and return predictions.
5. **Graph Algorithms for Social Network Analysis**

* Use NetworkX to analyze Twitter connections, LinkedIn graphs, or social networks.
* Implement BFS, DFS, and centrality measures.

**5 Advanced Projects**

These involve deeper model tuning, deployment, and real-world applications.

1. **Deep Learning for Image Classification (CNNs with TensorFlow/PyTorch)**

* Use a dataset like CIFAR-10 or MNIST.
* Train a Convolutional Neural Network (CNN).

1. **Recommender System (Collaborative or Content-Based Filtering)**

* Build a Movie Recommendation system (Netflix-style).
* Use Surprise Library or implement Matrix Factorization.

1. **Big Data Processing with PySpark**

* Work with a large dataset using Spark RDDs or DataFrames.
* Perform data aggregation and transformations.

1. **NLP Sentiment Analysis on Tweets or Reviews**

* Use NLTK, SpaCy, or Hugging Face transformers.
* Train a model to classify positive/negative sentiment.

1. **Full-Stack Data Science Project with Dashboard**

* Train an ML model and deploy it using Flask/FastAPI.
* Build a frontend dashboard with Streamlit or Dash.
* Deploy using Heroku or AWS.

**Final Steps for a Paid Internship**

* **Portfolio on GitHub:** Document all projects, showing code, results, and insights.
* **LinkedIn Profile & Resume:** Add project links, optimize keywords, and engage with DS content.
* **Kaggle & Leetcode Practice:** Participate in competitions and practice DSA/SQL queries.
* **Networking & Applications:** Apply for internships on LinkedIn, Glassdoor, and company websites.