KDnuggets

Subscribe to KDnuggets News | | **Contact**

search KDnuggets   [Search]

- SOFTWARE
- News/Blog
- Top stories
- Opinions
- Tutorials
- JOBS
- Companies
- Courses
- Datasets
- EDUCATION
- Certificates
- Meetings
- Webinars

PAW London, The Premier Machine Learning Conference, 17-18 October

KDnuggets Home » News » 2018 » Jun » Opinions, Interviews » 5 Data Science Projects That Will Get You Hired in 2018 (18:n25)

5 Data Science Projects That Will Get You Hired in 2018

◄ **Previous post**
Next post ►

[Like 269] [Share 269]   Tweet   [G+]   Share   **352**

Tags: Data Preparation, Data Science, Data Visualization, Hiring, Jupyter, Machine Learning

A portfolio of real-world projects is the best way to break into data science. This article highlights the 5 types of projects that will help land you a job and improve your career.

By John Sullivan, DataOptimal

You've been taking MOOCs and reading a bunch of textbooks, but now what do you do? Getting a job in data science can seem daunting. The best way to showcase your skills is with a portfolio. This shows employers that you can use the skills you've been learning.

To showcase these skills, here's 5 types of data science projects for your portfolio:

1. **Data Cleaning**

Data scientists can expect to spend up to 80% of the time on a new project cleaning data. This is a huge pain point for teams. If you can show that you're experienced at cleaning data, you'll immediately be more valuable. To create a data cleaning project, find some messy data sets, and start cleaning.

If you're working with Python, Pandas is a great library to use, and if you're working with R, you can use the dplyr package. Make sure to showcase the following skills:

- Importing data
- Joining multiple datasets
- Detecting missing values
- Detecting anomalies
- Imputing for missing values
- Data quality assurance

2. **Exploratory Data Analysis**

Another important aspect of data science is exploratory data analysis (EDA). This is the process of generating questions, and investigating them with visualizations. EDA allows an analyst to draw conclusions from data to drive business impact. It might include interesting insights based on customer segments, or sales trends based on seasonal effects. Often you can make interesting discoveries that weren't initial considerations.

Some useful Python libraries for exploratory analysis are [Pandas](#) and [Matplotlib](#). For R users, the [ggplot2](#) package will be useful. An EDA project should show the following skills:

- Ability to formulate relevant questions for investigation
- Identifying trends
- Identifying covariation between variables
- Communicating results effectively using visualizations (scatterplots, histograms, box and whisker, etc.)

3. **Interactive Data Visualizations**

Interactive data visualizations include tools such as dashboards. These tools are useful for both data science teams, as well as more business-oriented end users. Dashboards allow data science teams to collaborate, and draw insights together. Even more important, they provide an interactive tool for business-oriented customers. These individuals focus on strategic goals rather than technical details. Often the deliverable for a data science project to a client will be in the form of a dashboard.
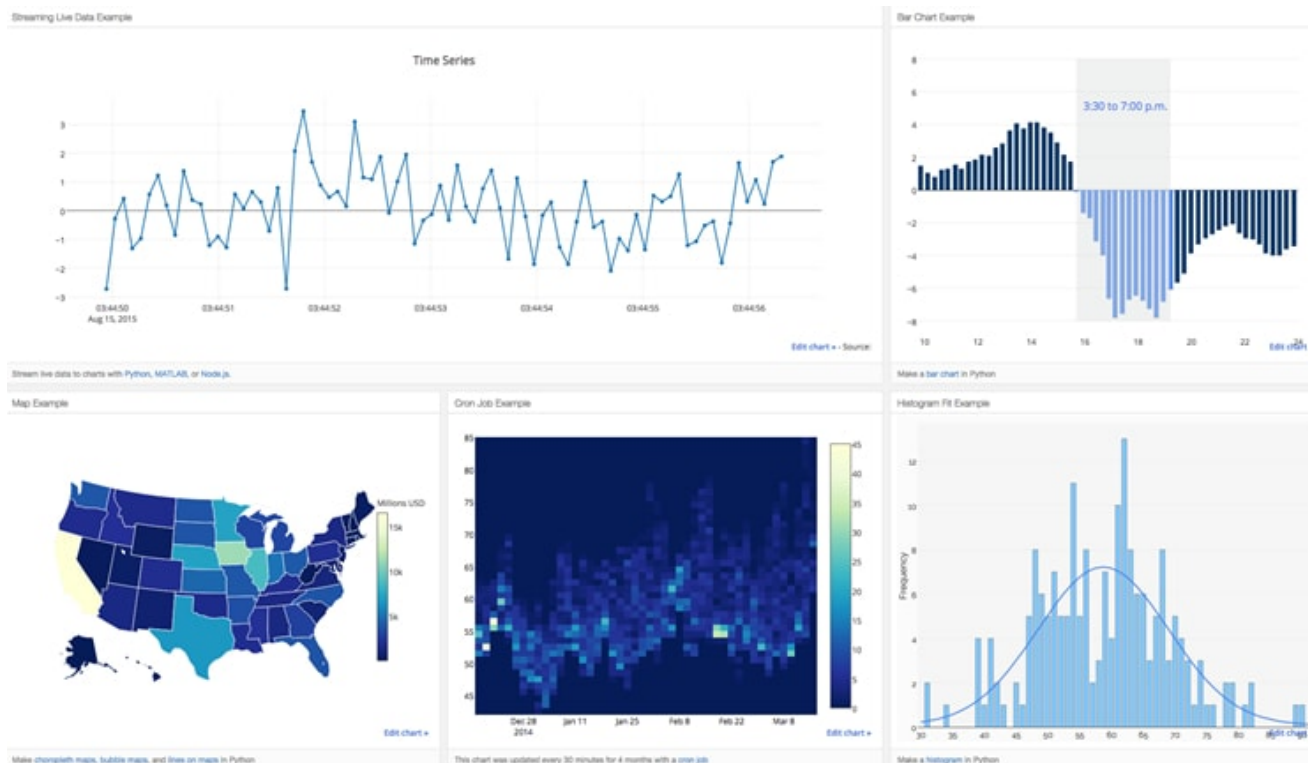
For Python users, the [Bokeh](#) and [Plotly](#) libraries are great for creating dashboards. For R users, be sure to check out RStudio's [Shiny](#) package. Your dashboard project should highlight these important skills:

- Including metrics relevant to your customer's needs
- Creating useful features
- A logical layout ("F-pattern" for easy scanning)
- Creating an optimum refresh rate
- Generating reports or other automated actions

4. **Machine Learning**

A machine learning project is another important piece of your data science portfolio. Now before you run off and start building some deep learning project, take a step back for a minute. Rather than building a complex machine learning model, stick with the basics. Linear regression and logistic regression are great to start with. These models are easier to interpret and communicate to upper level management. I'd also recommend focusing on a project that has a business impact, such as predicting customer churn, fraud detection, or loan default. These are more real-world than predicting flower type.
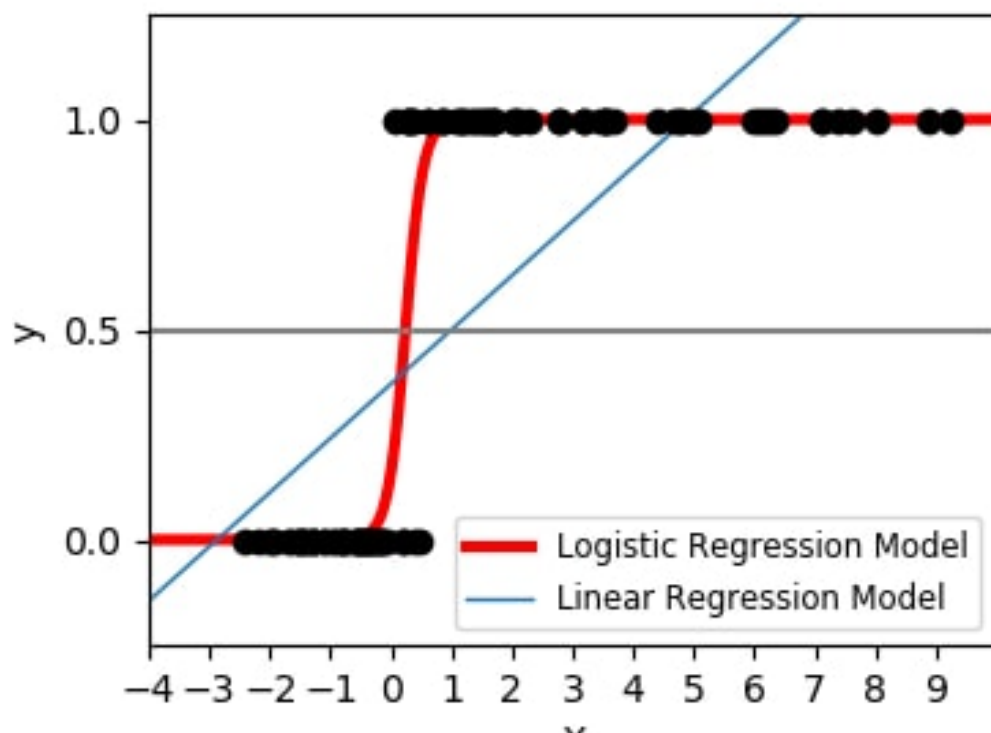
If you're a Python user, use the [Scikit-learn](#) library. For R users, use the [Caret](#) package. Your machine learning project should convey the following skills:

- Reason why you chose to use a specific machine learning model
- Splitting data into training/test sets (k-fold cross validation) to avoid overfitting
- Selecting the right evaluation metrics (AUC, adj-R^2, confusion matrix, etc.)
- Feature engineering and selection
- Hyperparameter tuning

5. **Communication**

Communication is an important aspect of data science. Effectively communicating results is what separates the good data scientists from the great ones. It doesn't matter how fancy your model is, if you can't explain it to teammates or customers, you won't get their buy-in. Slides and notebooks are both great communication tools. Use one of your machine learning projects and put it into slide format. You could also use a [Jupyter Notebook](#) or [RMarkdown](#) file for a communication project.
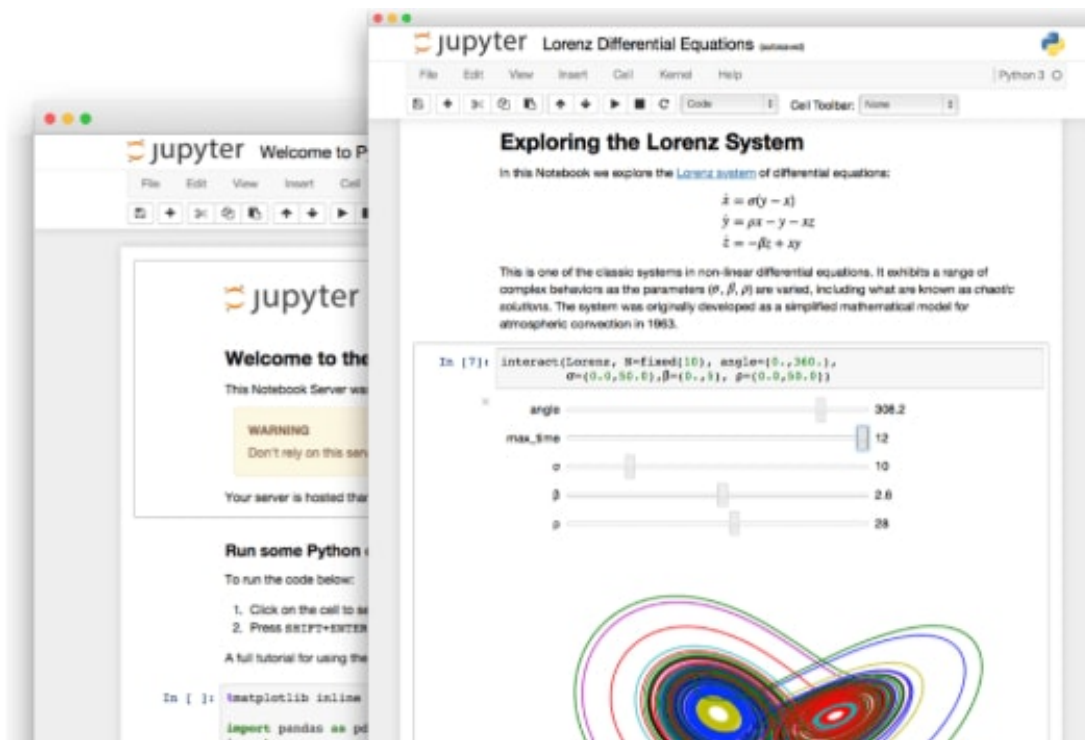
Make sure to understand who your intended audience is. Presenting to executives is very different than presenting to machine learning experts. Make sure to hit on these skills:

- Know your intended audience
- Present relevant visualizations
- Don't crowd your slides with too much information
- Make sure your presentation flows well
- Tie results to a business impact (reduced cost, increased revenue)

Make sure to document your projects in Jupyter Notebooks or RMarkdown files. You can then convert these markdown files to static websites for free using Github Pages. This is a great way to showcase your portfolio to potential employers.

Stay positive, continue to build projects, and you'll be on your way to landing a job in data science. Happy job hunting!

Bio: John Sullivan is the founder of the data science learning blog, DataOptimal. You can follow him on Twitter @DataOptimal.

Related:

- Data Science Predicting The Future
- Stagraph – a general purpose R GUI, for data import, wrangling, and visualization
- 7 Simple Data Visualizations You Should Know in R

11 Comments   KDnuggets   Mahmoud Danesh

♡

Important Update

When you log in with Disqus, we process personal data to facilitate your authentication and posting of comments. We also store the comments you post and those comments are immediately viewable and searchable by anyone around the world.

☐ I agree to Disqus' Terms of Service

☐ I agree to Disqus' processing of email and IP address, and the use of cookies, to facilitate my authentication and posting of comments, explained further in the Privacy Policy

Proceed

Gino Zhang • 3 months ago
how can i get the 5 project with detail py code??
1 ∧ | ∨ • Reply • Share ›

Yatin Ubhaykar • 2 months ago
#5 should be #1
1 ∧ | ∨ • Reply • Share ›

Gregory Piatetsky Mod ➜ Yatin Ubhaykar • 2 months ago
before you communicate (#5) you need to have some results to communicate, so you need to do some of the projects 1-4 before 5.
2 ∧ | ∨ • Reply • Share ›

Yatin Ubhaykar ➜ Gregory Piatetsky • 2 months ago
#1 not in terms of chronology, but in terms of importance
1 ∧ | ∨ • Reply • Share ›

NsikanAbasi Udo • 2 months ago
Thanks for sharing. It gonna help me organize my work as an upcoming Data Scientist.
∧ | ∨ • Reply • Share ›

KhuShi DhUpia • 2 months ago
can we get the data for smart grid analytics for research purpose?
∧ | ∨ • Reply • Share ›

Michael Muttiah • 3 months ago
Great article, thank you
∧ | ∨ • Reply • Share ›

mohit • 3 months ago
I am novice can you guide me, which projects should I start with @JohnSullivan
∧ | ∨ • Reply • Share ›

Archanous • 3 months ago
Great article!
∧ | ∨ • Reply • Share ›

DoNotLookDown • 3 months ago
Great tips! As an aspiring data scientist (of sorts), this is very useful.
∧ | ∨ • Reply • Share ›

✉ Subscribe   Ⓓ Add Disqus to your siteAdd DisqusAdd   🔒 Disqus' Privacy PolicyPrivacy PolicyPrivacy

◀ **Previous post**
Next post ▶

Top Stories Past 30 Days

<table>
<tr><td align="center">**Most Popular**</td><td align="center">**Most Shared**</td></tr>
</table>

Most Popular	Most Shared
1. Essential Math for Data Science: Why and How	1. Machine Learning Cheat Sheets
2. Machine Learning Cheat Sheets	2. Essential Math for Data Science: Why and How
3. Data Visualization Cheat Sheet	3. Journey to Machine Learning – 100 Days of ML Code
4. 5 Data Science Projects That Will Get You Hired in 2018	4. You Aren't So Smart: Cognitive Biases are Making Sure of It
5. Journey to Machine Learning – 100 Days of ML Code	5. Neural Networks and Deep Learning: A Textbook
6. Neural Networks and Deep Learning: A Textbook	6. 6 Steps To Write Any Machine Learning Algorithm From Scratch: Perceptron Case Study
7. 5 Resources to Inspire Your Next Data Science Project	7. A Winning Game Plan For Building Your Data Science Team

Latest News

- Colorado State University: Assistant Professor in Indus...
- Online Master's in Applied Data Science From Syracuse
- Basic Image Data Analysis Using Python – Part 4
- Why do I Call Myself a Data Scientist?
- A Concise Explanation of Learning Algorithms with the M...
- University of Nebraska at Omaha: Faculty Position in Co...

Penn State: Earn MS in Data Analytics online

**Top Stories
Last Week**

Most Popular

1. 🟢 NEW **Machine Learning Cheat Sheets**
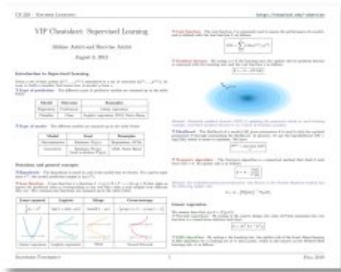
2. 🟢 NEW **Essential Math for Data Science: Why and How**
3. 🟢 NEW **6 Steps To Write Any Machine Learning Algorithm From Scratch: Perceptron Case Study**
4. 📈 **5 Data Science Projects That Will Get You Hired in 2018**
5. 🟢 NEW **How many data scientists are there and is there a shortage?**
6. 🟢 NEW **Unfolding Naive Bayes From Scratch**
7. 📈 **9 Must-have skills you need to become a Data Scientist, updated**

Most Shared

1. **Learning mathematics of Machine Learning: bridging the gap**

2. **Math for Machine Learning**
3. **Introducing Path Analysis Using R**
4. **Introduction to Deep Learning**
5. **Building a Machine Learning Model through Trial and Error**
6. **Deep Learning Framework Power Scores 2018**
7. **Unfolding Naive Bayes From Scratch**

KDnuggets Home » News » 2018 » Jun » Opinions, Interviews » 5 Data Science Projects That Will Get You Hired in 2018 (18:n25)