f (https://www.facebook.com/AnalyticsVidhya)   **🐦** (https://twitter.com/analyticsvidhya)

8+ (https://plus.google.com/+Analyticsvidhya/posts)

in (https://www.linkedin.com/groups/Analytics-Vidhya-Learn-everything-about-5057165)

☰

Analytics Vidhya (https://www.analyticsvidhya.com)
Learn Everything About Analytics

Home (https://www.analyticsvidhya.com/) › Machine Learning (https://www.analyticsvidhya.com/blog/category/machine-learning...
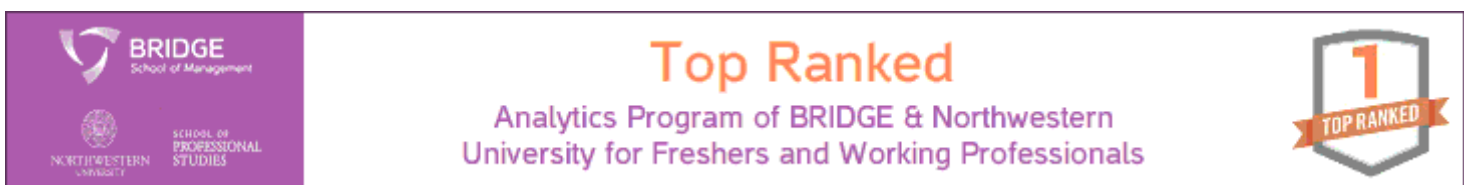
17 Ultimate Data Science Projects To Boost Your Knowledge and Skills (& can be accessed freely)

MACHINE LEARNING (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/MACHINE-LEARNING/)   PYTHON

(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/PYTHON-2/)   R (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/)

csvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-

Boost%20Your%20Knowledge%20and%20Skills%20(&%20can%20be%20accessed%20freely)) **🐦** (https://twitter.com/home?

st%20Your%20Knowledge%20and%20Skills%20(&%20can%20be%20accessed%20freely)+https://www.analyticsvidhya.com/blog/2016/10/17-

(https://plus.google.com/share?url=https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-

ww.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-

16/10/17-DATA-SCIENCE-

jects%20To%20Boost%20Your%20Knowledge%20and%20Skills%20(&%20can%20be%20accessed%20freely))

BRIDGE School of Management

NORTHWESTERN UNIVERSITY SCHOOL OF PROFESSIONAL STUDIES

Top Ranked
Analytics Program of BRIDGE & Northwestern
University for Freshers and Working Professionals

1 TOP RANKED

(http://admissions.bridgesom.com/pba-new/?
utm_source=AV&utm_medium=BannerInline&utm_campaign=AVBanner20August)

Introduction

Data science projects offer you a promising way to kick-start your analytics career. Not only you get to learn data science by applying, you also get projects to showcase on your CV. Nowadays, recruiters evaluate a candidate's potential by his/her work, not as much by certificates and resumes. It wouldn't matter, if you just tell them how much you know, if you have nothing to show them! That's where most people struggle and miss out!

You might have worked on several problems, but if you can't make it presentable & explanatory, how on earth would someone know what you are capable of? That's where these projects would help you. Think of the time spend on these projects like your training sessions. I guarantee, the more time you spend, the better you'll become!

The data sets in the list below are handpicked. I've made sure to provide you a taste of variety of problems from different domains with different sizes. I believe, everyone must learn to smartly work on large data sets, hence large data sets are added. Also, I've made sure all the data sets are open and free to access.
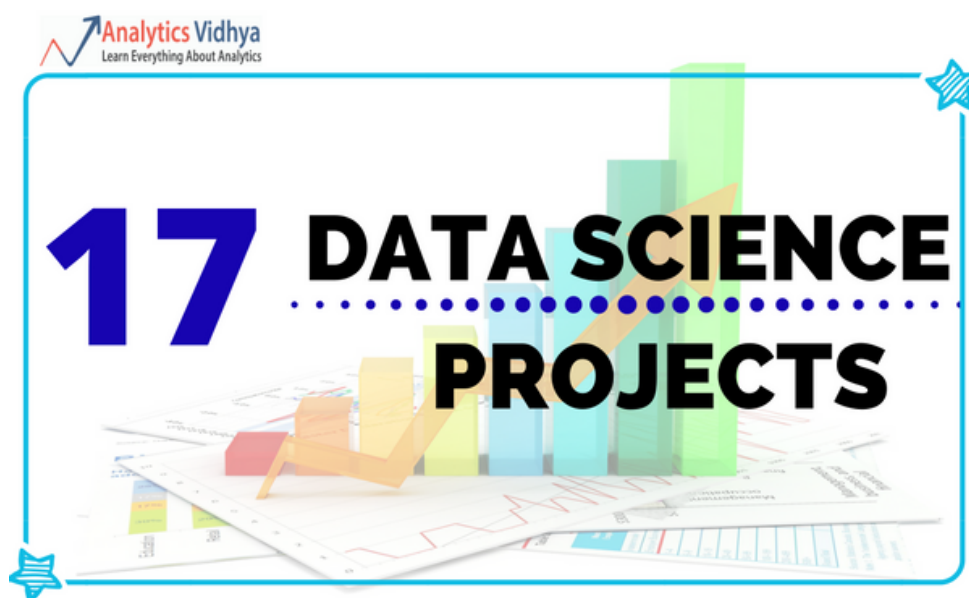
Useful Information

To help you decide your start line, I've divided the data set into 3 levels namely:

1. **Beginner Level:** This level comprises of data sets which are fairly easy to work with, and doesn't require complex data science techniques. You can solve them using basic regression / classification algorithms. Also, these data sets have enough open tutorials to get you going. In this list, I've provided tutorials also to help you get started.
2. **Intermediate Level:** This level comprises of data sets which are challenging. It consists of mid & large data sets which require some serious pattern recognition skills. Also, feature engineering will make a difference here. There is no limit of use of ML techniques, everything under the sun can be put to use.
3. **Advanced Level:** This level is best suited for people who understand advanced topics like neural networks, deep learning, recommender systems etc. High dimensional data are also featured here. Also, this is the time to get creative – see the creativity best data scientists bring in their work and codes.

Table of Contents

Beginner Level

1. Iris Data Set

This is probably the most versatile, easy and resourceful data set in pattern recognition literature. Nothing could be simpler than iris data set to learn classification techniques. If you are totally new to data science, this is your start line. The data has only 150 rows & 4 columns.

Problem: Predict the flower class based on available attributes.

Start: Get Data (https://archive.ics.uci.edu/ml/datasets/Iris) | **Tutorial:** Get Here (http://www.slideshare.net/thoi_gian/iris-data-analysis-with-r)
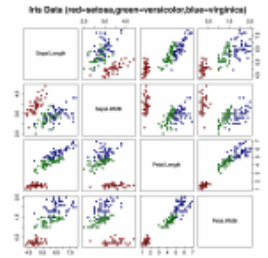
2. Titanic Data Set

This is another most quoted data set in global data science community. With several tutorials and help guides, this project should give you enough kick to pursue data science deeper. With healthy mix of variables comprising categories, numbers, text, this data set has enough scope to support crazy ideas! This is a classification problem. The data has 891 rows & 12 columns.

Problem: Predict the survival of passengers in Titanic.

Start: Get Data (https://www.kaggle.com/c/titanic) | **Tutorial:** Get Here (http://trevorstephens.com/kaggle-titanic-tutorial/getting-started-with-r/)

3. Loan Prediction Data Set

Among all industries, insurance domain has the largest use of analytics & data science methods. This data set would provide you enough taste of working on data sets from insurance companies, what challenges are faced, what strategies are used, which variables influence the outcome etc. This is a classification problem. The data has 615 rows and 13 columns.

Problem: Predict if a loan will get approved or not.

Start: Get Data (https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii/) | **Tutorial:** Get Here (https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/)

4. Bigmart Sales Data Set

Retail is another industry which extensively uses analytics to optimize business processes. Tasks like product placement, inventory management, customized offers, product bundling etc are being smartly handled using data science techniques. As the name suggests, this data comprises of transaction record of a sales store. This is a regression problem. The data has 8523 rows of 12 variables.

Problem: Predict the sales.

Start: Get Data (https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/) | **Tutorial:** Get Here (https://www.analyticsvidhya.com/blog/2016/02/bigmart-sales-solution-top-20/)

5. Boston Housing Data Set

This is another popular data set used in pattern recognition literature. The data set comes from real estate industry in Boston (US). This is a regression problem. The data has 506 rows and 14 columns. Thus, it's a fairly small data set where you can attempt any technique without worrying about your laptop's memory issue.

Problem: Predict the median value of owner occupied homes

Start: Get Data (http://archive.ics.uci.edu/ml/datasets/Housing) | **Tutorial:** Get Here (https://www.analyticsvidhya.com/blog/2015/11/started-machine-learning-ms-excel-xl-miner/)

Intermediate Level

1. Human Activity Recognition

This data set is collected from recordings of 30 human subjects captured via smartphones enabled with embedded inertial sensors. Many machine learning courses use this data for students practice. It's your turn now. This is a multi-classification problem. The data set has 10299 rows and 561 columns.

Problem: Predict the activity category of a human

Start: Get Data (http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones)

2. Black Friday Data Set

This data set comprises of sales transactions captured at a retail store. It's a classic data set to explore your feature engineering skills and day to day understanding from your shopping experience. It's a regression problem. The data set has 550069 rows and 12 columns.

Problem: Predict purchase amount.

Start: Get Data (https://datahack.analyticsvidhya.com/contest/black-friday/)

3. Text Mining Data Set

This data set is originally from siam competition 2007. The data set comprises of aviation safety reports describing problem(s) which occurred in certain flights. It is a multi-classification, high dimensional problem. It has 21519 rows and 30438 columns.

Problem: Classify the documents according to their labels

Start: Get Data (http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multilabel.html#siam-competition2007) | Get Information (https://catalog.data.gov/dataset/siam-2007-text-mining-competition-dataset/resource/794f14ae-8135-41d2-88c8-86bf8fad9cf6/proxy)

4. Trip History Data Set

This data set comes from a bike sharing service in US. This data set requires you to exercise your pro data munging skills. The data set is provided quarter wise from 2010 (Q4) onwards. Each file has 7 columns. It is a classification problem.

Problem: Predict the class of user

Start: Get Data (https://www.capitalbikeshare.com/trip-history-data)

5. Million Song Data Set

Didn't you know analytics can be used in entertainment industry also? Do it yourself now. This data set puts forward a regression task. It consists of 515345 observations and 90 variables. However, this is just a tiny subset of original database (http://labrosa.ee.columbia.edu/millionsong/pages/getting-dataset) of million song data. You should use data linked below.

Problem: Predict release year of the song

Start: Get Data (http://archive.ics.uci.edu/ml/datasets/YearPredictionMSD)

6. Census Income Data Set

It's an imbalanced classification and a classic machine learning problem. You know, machine learning is being extensively used to solve imbalanced problems such as cancer detection, fraud detection etc. It's time to get your hand dirty. The data set has 48842 rows and 14 columns. For guidance, you

can   check   my   imbalanced   data   project
(https://www.analyticsvidhya.com/blog/2016/09/this-machine-learning-
project-on-imbalanced-data-can-add-value-to-your-resume/).

Problem: Predict the income class of US population

Start: Get Data (http://archive.ics.uci.edu/ml/machine-learning-databases/census-income-mld/)

7. Movie Lens Data Set

This data set allows you to build a recommendation engine. Have you created one before? It's one of the most popular & quoted data set in data science industry. It is available in various dimensions (http://grouplens.org/datasets/movielens/). Here I've used a fairly small size. It has 1 million ratings from 6000 users on 4000 movies.
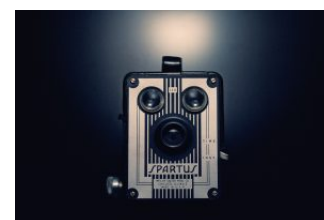
Problem: Recommend new movies to users

Start: Get Data (http://grouplens.org/datasets/movielens/1m/)

Advanced Level

1. Identify your Digits Data Set

This data set allows you to study, analyze and recognize elements in the images. That's exactly how your camera detects your face, using image recognition! It's your turn to build and test that technique. It's an digit recognition problem. This data set has 7000 images of 28 X 28 size, sizing 31MB.

Problem: Identify digits from an image

Start: Get Data (https://datahack.analyticsvidhya.com/contest/practice-problem-identify-the-digits/)

2. Yelp Data Set

This data set is a part of round 8 of The Yelp Dataset Challenge. It comprises of nearly 200,000 images, provided in 3 json files of ~2GB. These images provide information about local businesses in 10 cities across 4 countries. You are required to find insights from data using cultural trends, seasonal trends, infer categories, text mining, social graph mining etc.

Problem: Find insights from images

Start: Get Data (https://www.yelp.com/dataset_challenge)

3. Image Net Data Set

ImageNet offers variety of problems which encompasses object detection, localization, classification and screen parsing. All the images are freely available. You can search for any type of image and build your project around it. As of now, this image engine has 14,197,122 images of multiple shapes sizing up to 140GB.
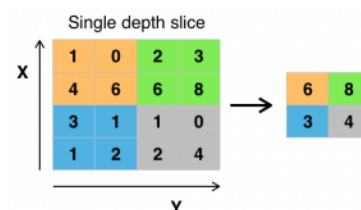
Problem: Problem to solve is subjected to the image type you download

Start: Get Data (http://image-net.org/download-imageurls)

4. KDD 1999 Data Set

How could I miss KDD Cup? Originally, KDD brought the taste of data mining competition to the world. Don't you want to see what data set they used to offer? I assure you, it'll be an enriching experience. This data poses a classification problem. It has 4M rows and 48 columns in a ~1.2GB file.
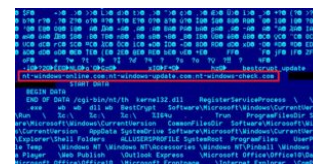
Problem: Classify a network intrusion detector as good or bad.

Start: Get Data (https://archive.ics.uci.edu/ml/datasets/KDD+Cup+1999+Data)

5. Chicago Crime Data Set

The ability of handle large data sets is expected of every data scientist these days. Companies no longer prefer to work on samples, they use full data. This data set would provide you much needed hands on experience of handling large data sets on your local machines. The problem is easy, but data management is the key! This data set has 6M observations. It's a multi-classification problem.

Problem: Predict the type of crime.

Start: Get Data (https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2) | To download data, click on Export -> CSV

End Notes

Out of the 17 data sets listed above, you should start by finding the right match of your skills. Say, if you are a beginner in machine learning, avoid taking up advanced level data sets. Don't bite more than you can chew and don't feel overwhelmed with how much you still have to do. Instead, focus on making step wise progress.

Once you complete 2 – 3 projects, showcase them on your resume and your github profile (most important!). Lots of recruiters these days hire candidates by stalking github profiles. Your motive shouldn't be to do all the projects, but to pick out selected ones based on data set, domain, data set size whichever excites you the most. If you want me to solve any of above problem and create a complete project like this (https://www.analyticsvidhya.com/blog/2016/09/this-machine-learning-project-on-imbalanced-data-can-add-value-to-your-resume/), let me know.

Did you find this article useful ? Have you already build any project on these data sets? Do share your experience, learnings and suggestions in comments below.

You can test your skills and knowledge. Check out Live Competitions (http://datahack.analyticsvidhya.com/contest/all) and compete with best Data Scientists from all over the world.

Share this:

in (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=linkedin&nb=1)

142

f (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=facebook&nb=1)

G+ (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=google-plus-1&nb=1)

🐦 (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=twitter&nb=1)

📖 (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=pocket&nb=1)

🔴 (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/?share=reddit&nb=1)

RELATED

(https://www.analyticsvidhya.com/blog/2014/11/data-science-projects-learn/)
Five data science projects to learn data science (https://www.analyticsvidhya.com/blog/2014/11/data-science-projects-learn/)
In "Big data"

(https://www.analyticsvidhya.com/blog/2015/07/github-special-data-scientists-to-follow-best-tutorials/)
Top Data Scientists to Follow & Best Data Science Tutorials on GitHub (https://www.analyticsvidhya.com/blog/2015/07/github-special-data-scientists-to-follow-best-tutorials/)
In "Business Analytics"
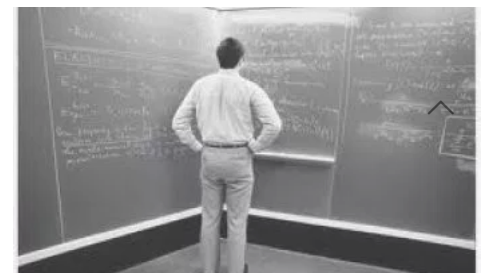
(https://www.analyticsvidhya.com/blog/2014/11/data-science-manager-lead-team/)
5 things every data science manager should do before leading a team (https://www.analyticsvidhya.com/blog/2014/11/data-science-manager-lead-team/)
In "Big data"

TAGS: ANALYTICS PROJECTS (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ANALYTICS-PROJECTS/), BOSTON HOUSING DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BOSTON-HOUSING-DATA/), CLASSIFICATION DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/CLASSIFICATION-DATA/), DATA EXPLORATION (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DATA-EXPLORATION/), DATA MINING PROJECTS (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DATA-MINING-PROJECTS/), DATA SCIENCE PROJECT (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DATA-SCIENCE-PROJECT/), FEATURE ENGINEERING (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/FEATURE-ENGINEERING/), IMAGE RECOGNITION (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/IMAGE-RECOGNITION/), MACHINE LEARNING (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MACHINE-LEARNING/), MACHINE LEARNING PROJECTS (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MACHINE-LEARNING-PROJECTS/), MNIST DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MNIST-DATA/), MULTI-CLASSIFICATION (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MULTI-CLASSIFICATION/), OBJECT RECOGNITION (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/OBJECT-RECOGNITION/), REGRESSION DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REGRESSION-DATA/), TEXT MINING DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/TEXT-MINING-DATA/), TITANIC DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/TITANIC-DATA/), YELP DATA (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/YELP-DATA/)

Previous Article

Complete Guide on DataFrame Operations in PySpark (https://www.analyticsvidhya.com/blog/2016/10/spark-dataframe-and-operations/)

(https://www.analyticsvidhya.com/blog/author/manish-saraswat/)

Author

Manish Saraswat (https://www.analyticsvidhya.com/blog/author/manish-saraswat/) ︿

I believe education can change this world. Knowledge is the most powerful asset one should own. I am passionate about helping people. I love solving problems. I care about animals, unprivileged people, knowledge, health and books. I use R to build ML models. data.table, ggplot, h2o, mlr are my best friends.

in (https://in.linkedin.com/in/saraswatmanish)

14 COMMENTS

Demudu says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?

REPLYTOCOM=117503#RESPOND)2016 AT 6:02 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117503)

Fantastic

Manish Saraswat says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117512#RESPOND)
OCTOBER 26, 2016 AT 8:56 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117512)

Thank you! 🙂

Tesfaye says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117536#RESPOND) OCTOBER 26, 2016 AT 7:41 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117536)

Thank you Manish. These are wonderful project ideas. Please I would love to have sample
solutions if/when you have them.

Mallikarjun (https://in.linkedin.com/in/mrvelameli) says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117505#RESPOND) OCTOBER 26, 2016 AT 6:31 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117505)

Thank You so much… 🙂 I have been wondering, how to start with projects. This will help me out. I
have done machine learning course of Prof. Andrew Ng. and I have good knowledge of statistics
and R and Matlab. Please let me know, if any skill required to be a data scientist. Thank you again.
🙂

Manish Saraswat says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117511#RESPOND)
OCTOBER 26, 2016 AT 8:55 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117511)

Hi Mallikarjun, I received several emails and messages to help people in selecting their data
science projects, which motivated me to write this post. I'd suggest you to take up any project
according to your understanding and start working on it. Through the way you'll discover topics
which you are yet to pick up. All the best!

venugopal rao says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117506#RESPOND)

OCTOBER 26, 2016 AT 7:18 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117506)

Great Collection.. All together in one place

Manish Saraswat says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117510#RESPOND)

OCTOBER 26, 2016 AT 8:52 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117510)

Thanks :)

Akash says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117515#RESPOND)

OCTOBER 26, 2016 AT 10:17 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117515)

Hey Manish,

Good one. Do you have the 'R' equivalent for Boston housing and Big mart sales ?

Thanks,

Femy (http://www.6tides.com) says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117521#RESPOND)

OCTOBER 26, 2016 AT 12:29 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117521)

Do you have anything on operational risk or risk in general especially consumer credit risk

---

Manish Saraswat says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?REPLYTOCOM=117532#RESPOND)

OCTOBER 26, 2016 AT 5:06 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117532)

Hi Femy, there are several data sets I came across while composing this post. For your persual, I would suggest you to look at https://github.com/caesar0301/awesome-public-datasets (https://github.com/caesar0301/awesome-public-datasets)
Chances are you'll find the data you are looking for!

Sridevi says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117524#RESPOND) OCTOBER 26, 2016 AT 12:54 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117524)

Hi Manish,

Wonderful collection. Great resource for exploring ML.

Thanks

Krishna says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117526#RESPOND) OCTOBER 26, 2016 AT 1:02 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117526)

Hi Manish,

Can you please give some insights about the Knoctober data which was conducted recently?

Manish Saraswat says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117531#RESPOND) OCTOBER 26, 2016 AT 5:03 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117531)

Hi Krishna, tomorrow we'll post a complete article on knocktober competition so that people like you can gather their learnings. Do keep a check on your in mails.

Tesfaye says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-BOOST-YOUR-KNOWLEDGE-AND-SKILLS/?
REPLYTOCOM=117537#RESPOND) OCTOBER 26, 2016 AT 7:42 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/10/17-ULTIMATE-DATA-SCIENCE-PROJECTS-TO-
BOOST-YOUR-KNOWLEDGE-AND-SKILLS/#COMMENT-117537)

Thank you Manish. These are wonderful project ideas. Please I would love to have sample solutions if/when you have them.

LEAVE A REPLY

Connect with:

[f] (https://www.analyticsvidhya.com/wp-login.php?
action=wordpress_social_authenticate&mode=login&provider=Facebook&redirect_to=https%3A%2F%2Fwww.ar
ultimate-data-science-projects-to-boost-your-knowledge-and-skills%2F)

Your email address will not be published.

Comment

Name (required)

Email (required)

Website ⌃

SUBMIT COMMENT

TOP AV USERS

Rank	Name	Points
1	SRK (https://datahack.analyticsvidhya.com/user/profile/SRK)	5388
2	Aayushmnit (https://datahack.analyticsvidhya.com/user/profile/aayushmnit)	4928

3		Nalin Pasricha (https://datahack.analyticsvidhya.com/user/profile/Nalin)	4417
4		vopani (https://datahack.analyticsvidhya.com/user/profile/Rohan Rao)	4403
5		binga (https://datahack.analyticsvidhya.com/user/profile/binga)	3371

More Rankings (http://datahack.analyticsvidhya.com/users)

POPULAR POSTS

ᐱ

- A Complete Tutorial to Learn Data Science with Python from Scratch
  (https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-
  2/)
- A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)
  (https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-
  python/)
- Essentials of Machine Learning Algorithms (with Python and R Codes)
  (https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/)
- Most Active Data Scientists, Free Books, Notebooks & Tutorials on Github
  (https://www.analyticsvidhya.com/blog/2016/09/most-active-data-scientists-free-books-notebooks-
  tutorials-on-github/)
- 7 Types of Regression Techniques you should know!
  (https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/)

- 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)
  (https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/)
- Beginner's guide to Web Scraping in Python (using BeautifulSoup)
  (https://www.analyticsvidhya.com/blog/2015/10/beginner-guide-web-scraping-beautiful-soup-python/)
- A Complete Tutorial on Time Series Modeling in R
  (https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/)

ENTER THE EXCITING FIELD OF ANALYTICS
LEARN TOOLS SUCH AS
R HADOOP
PYTHON BIG DATA
www.imarticus.org

(http://imarticus.org/diploma-in-big-data-analytics?
id=AnalyticsVidhya)

RECENT POSTS

17 Ultimate Data Science Projects To Boost Your Knowledge and Skills (& can be accessed freely)

(https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/)

(https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/)

MANISH SARASWAT , OCTOBER 26, 2016

(https://www.analyticsvidhya.com/blog/2016/10/spark-dataframe-and-operations/) **Complete Guide on DataFram**

e Operations in PySpark (https://www.analyticsvidhya.com/blog/2016/10/spark-dataframe-and-operations/)

ANKIT GUPTA , OCTOBER 23, 2016

(https://www.analyticsvidhya.com/blog/2016/10/winning-strategies-for-ml-competitions-from-past-winners/)

Winning Strategies for ML Competitions from Past Winners (https://www.analyticsvidhya.com/blog/2016/10/winning-strategies-for-ml-competitions-from-past-winners/)

KUNAL JAIN , OCTOBER 21, 2016

(https://www.analyticsvidhya.com/blog/2016/10/16-new-must-watch-tutorials-courses-on-machine-learning/)

16 New Must Watch Tutorials, Courses on Machine Learning (https://www.analyticsvidhya.com/blog/2016/10/16-new-must-watch-tutorials-courses-on-machine-learning/)

MANISH SARASWAT , OCTOBER 19, 2016

GET CONNECTED

6,837
FOLLOWERS
(http://www.twitter.com/analyticsvidhya)

21,724
FOLLOWERS
(http://www.facebook.com/Analyticsvidhya)

g+ **1,376**
FOLLOWERS
(https://plus.google.com/+Analyticsvidhya)

Email
SUBSCRIBE
(http://feedburner.google.com/fb/a/mailverify?
uri=analyticsvidhya)

ABOUT US

(https://discuss.analyticsvidhya.com)
For those of you, who are wondering what is "Analytics Vidhya", "Analytics" can be defined as the science of
extracting insights from raw data. The spectrum of analytics starts from capturing data and evolves into using
insights / trends from this data to make informed decisions. Read More
(http://www.analyticsvidhya.com/about-me/)

STAY CONNECTED

6,837
FOLLOWERS
(http://www.twitter.com/analyticsvidhya)

21,724
FOLLOWERS
(http://www.facebook.com/Analyticsvidhya)

g+ # 1,376
FOLLOWERS
(https://plus.google.com/+Analyticsvidhya)

Email
SUBSCRIBE
(https://feedburner.google.com/fb/a/mailverify?
uri=analyticsvidhya)

LATEST POSTS

(https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-

boost-your-knowledge-and-skills/)

17Ultimate Data Science Projects To Boost Your Knowledge and Skills (& can be accessed freely) (https://www.analyticsvidhya.com/blog/2016/10/17-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/)
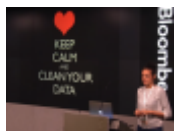
MANISH SARASWAT , OCTOBER 26, 2016

(https://www.analyticsvidhya.com/blog/2016/10/spark-dataframe-and-operations/) **Complete Guide on DataFram**

e Operations in PySpark (https://www.analyticsvidhya.com/blog/2016/10/spark-dataframe-and-operations/)

ANKIT GUPTA , OCTOBER 23, 2016

(https://www.analyticsvidhya.com/blog/2016/10/winning-strategies-for-ml-competitions-

from-past-winners/)

Winning Strategies for ML Competitions from Past Winners (https://www.analyticsvidhya.com/blog/2016/10/winning-strategies-for-ml-competitions-from-past-winners/)

KUNAL JAIN , OCTOBER 21, 2016

(https://www.analyticsvidhya.com/blog/2016/10/16-new-must-watch-tutorials-courses-on-

machine-learning/)

16 New Must Watch Tutorials, Courses on Machine Learning (https://www.analyticsvidhya.com/blog/2016/10/16-new-must-watch-tutorials-courses-on-machine-learning/)

MANISH SARASWAT , OCTOBER 19, 2016

QUICK LINKS

Home (https://www.analyticsvidhya.com/)

About Us (https://www.analyticsvidhya.com/about-me/)

Our team (https://www.analyticsvidhya.com/about-me/team/)

Privacy Policy (https://www.analyticsvidhya.com/privacy-policy/)

Refund Policy (https://www.analyticsvidhya.com/refund-policy/)

Terms of Use (https://www.analyticsvidhya.com/terms/)

TOP REVIEWS