Structuring your
1st Data Science Project

Date : 11-10-2020 | Speaker : Ayon Roy | Event : PSI Hacks

Visit - AYONROY.ML

Hello **Buddy!**

I am **Ayon Roy**

B.Tech CSE (2017-2021)

Data Science Intern @ **Lulu International Exchange**, Abu Dhabi

(**World's Leading Financial Services Company**)

Brought **Kaggle Days Meetup** Community in India for the 1st time

If you haven't heard about me yet, you might have been living under the rocks. Wake up !!

Agenda (11-10-2020)

- What is Data Science?
- What are the applications of Data Science?
- What are the steps to start Data Science?
- What are things to focus on while making a Data Science project?
- How to organize the structure of your 1st Data Science project?
- Why organizing a Data Science project matters?

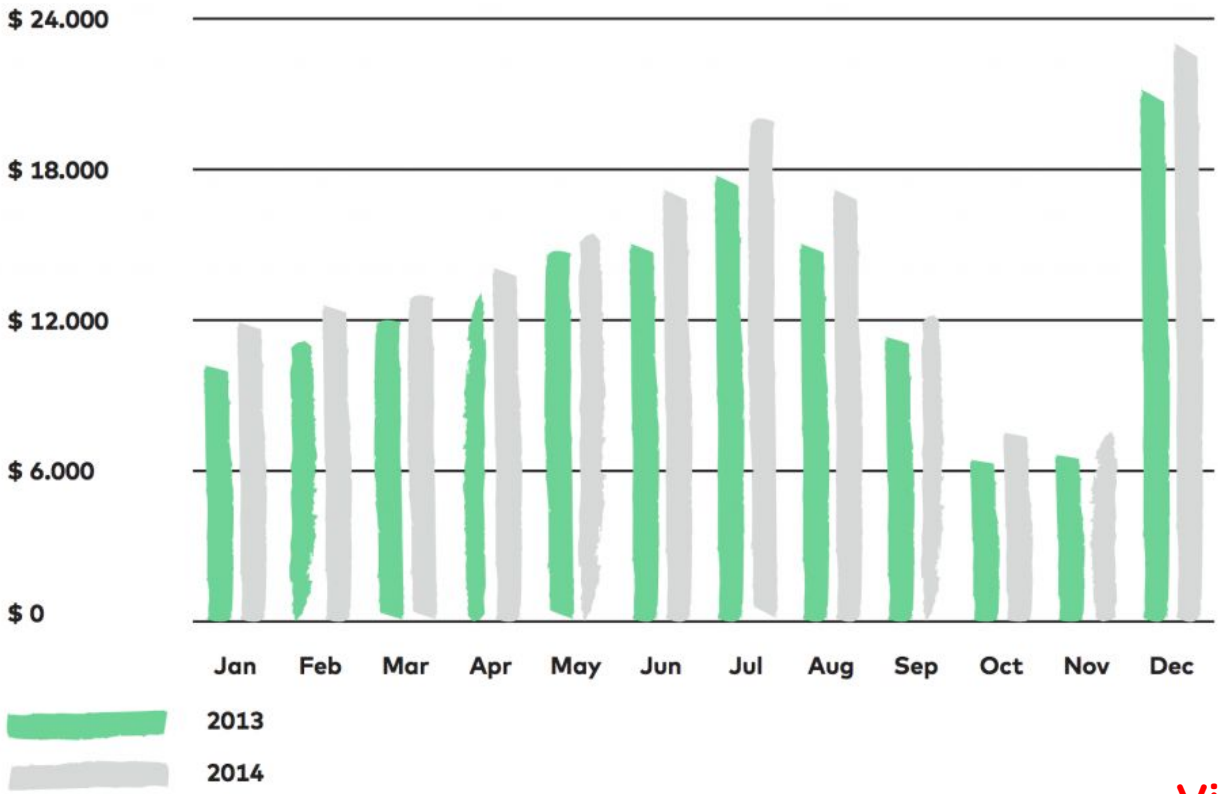
What is Data Science?

You have a large amount of data and you're trying to extract something smart and useful from it; here comes the power of Data Science.

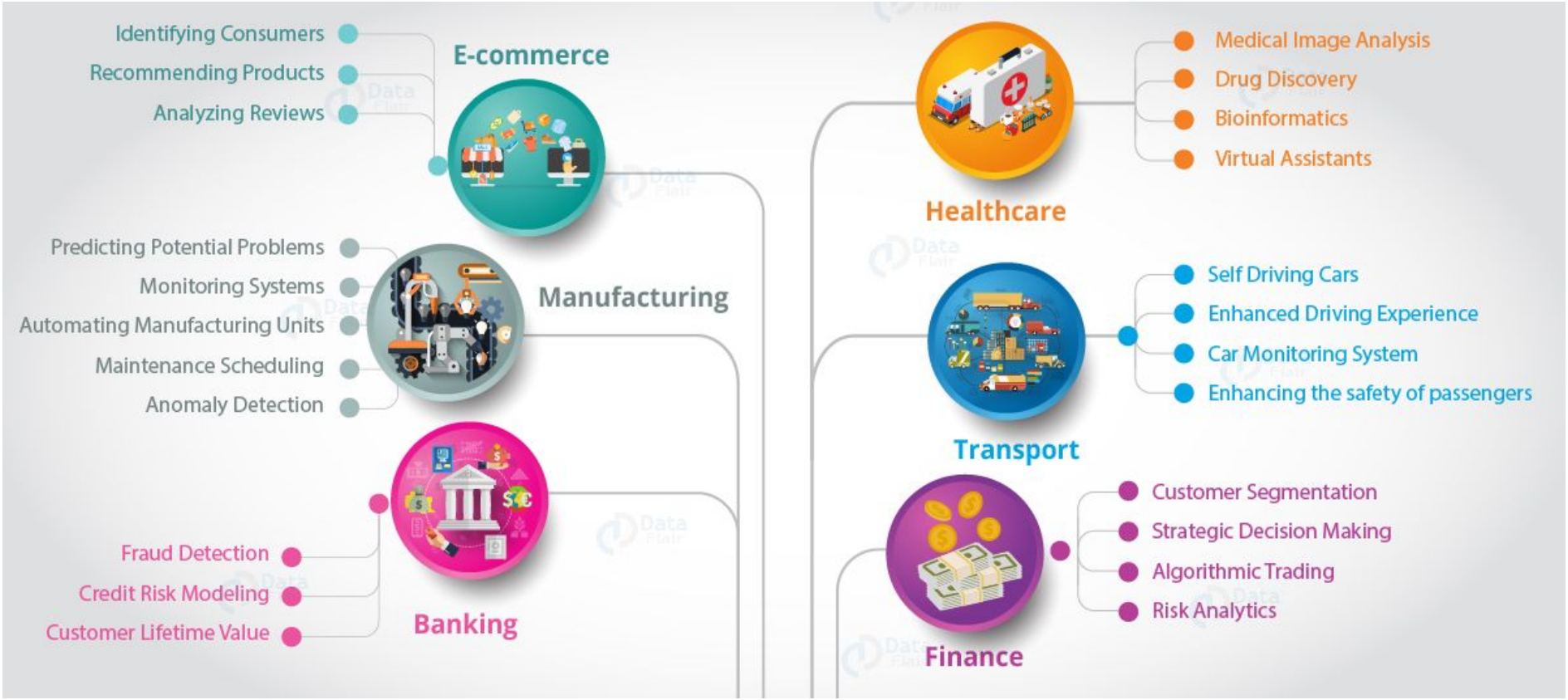
This is done with a combination of scientific disciplines. Like Mathematics, Statistics, Computer Science, etc (though they may not be an expert in all these fields).

Using Data Science, you can present the data in a much more useful form as compared to the raw data available initially from structured as well as unstructured forms.

Let's see how we can solve an E-Commerce Problem statement using Data Science



Applications of Data Science



Steps to learn

Data Science

Start with

Maths for Data Science

Visit - AYONROY.ML

But why should I do Maths first for Data Science?

- Week 1 : Linear Algebra [B] https://www.khanacademy.org/math/linear-algebra
- Week 2 : Calculus [B] https://www.youtube.com/playlist?list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr or https://www.mathsisfun.com/calculus/ ; want theoretical notes , find it at https://the-learning-machine.com/article/machine-learning/calculus .
- Week 3 : Probability [B] https://www.edx.org/course/introduction-probability-science-mitx-6-041x-2
- Week 4 : Statistics [B] http://alex.smola.org/teaching/cmu2013-10-701/stats.html
- Algorithms (Only if you want to learn proper software development) [Highly optional]
  This is an overview of what the students study as the subject Data Structures & Algorithm . So if you are fluent with this part , you can skip this !! https://www.edx.org/course/algorithm-design-analysis-pennx-sd3x

Start with Python

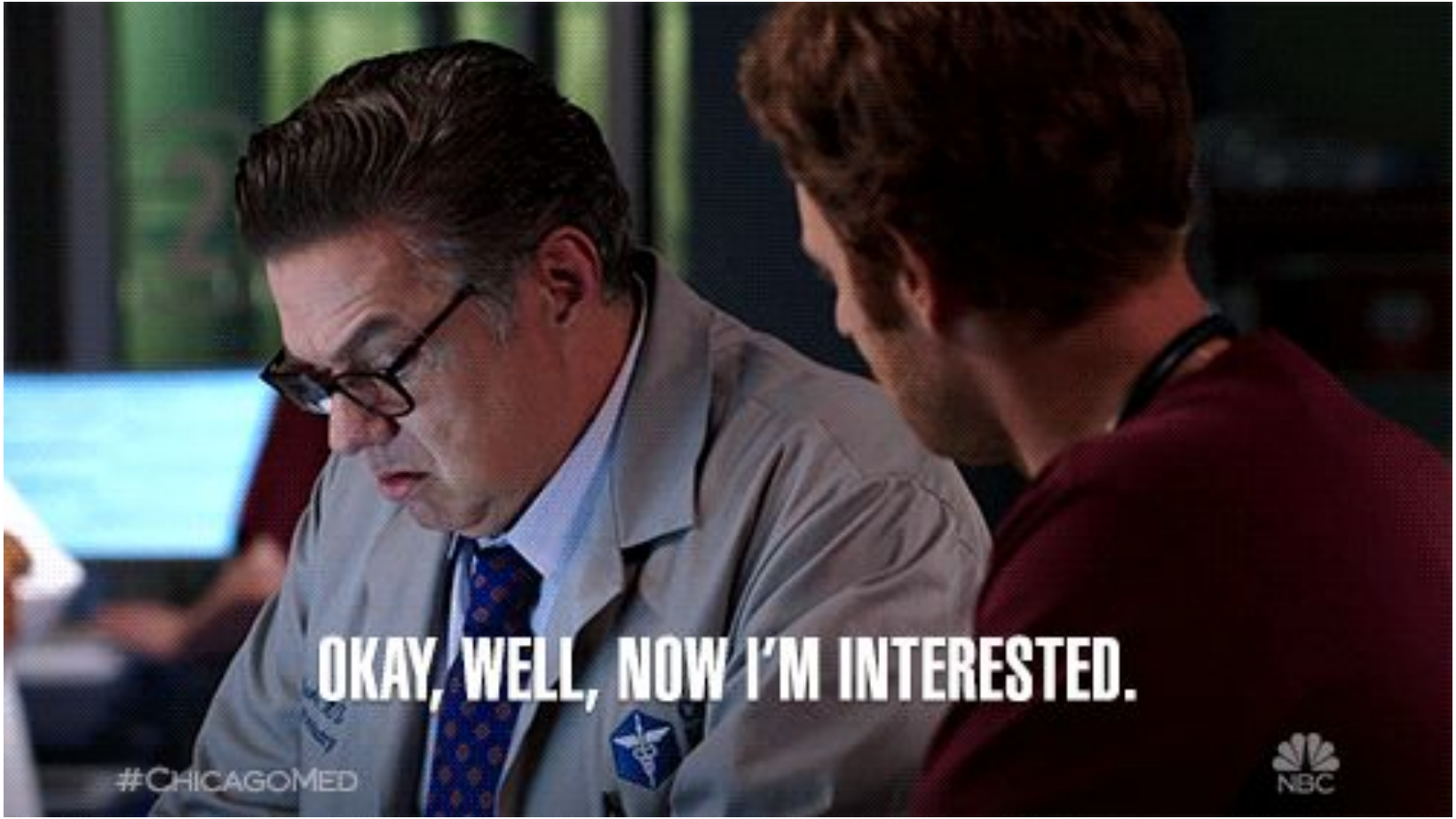
&

try to implement those

Mathematical Concepts

Start exploring Libraries
& then start Data Science,
Machine Learning Courses

Visit - AYONROY.ML



- Introduction to python for data science [B] https://www.datacamp.com/courses/intro-to-python-for-data-science
- Want to dive deeper into Data Visualization & Pre-Processing ? Look into Data Visualization & Pre-Processing section in miscellaneous resources . [Highly optional]
- Want to explore the field of Deep Learning ? See the Deep Learning Section in miscellaneous resources . [Highly optional]
- Want to explore the field of Natural Language Processing [NLP } ? See the Natural language Processing Section in miscellaneous resources . [Highly optional]
- See how ML codes are written and made to work at - > https://github.com/maykulkarni/Machine-Learning-Notebooks or https://github.com/GokuMohandas/practicalAI/blob/master/README.md . [Highly optional]
- Find useful resources here at https://github.com/ujjwalkarn/Machine-Learning-Tutorials/blob/master/README.md . [Highly optional]

Don’t rush behind completing Courses & add them to Resume

Understand the concepts well before starting Projects

Things to focus on while making a Data Science Project

Analytics Project Life Cycle

The 5 Phases

UNDERSTAND & PLAN

Translate the business question to analytics question

DATA

Work cross functionally to gather and process data

ANALYSIS

Solve problems

PRESENTATION

Present results and tell a story

DOCUMENTATION & REFLECTION

Make your project a happy ending

How to organize your 1st Data Science Project?

Local Project Directory	Github Repository

▪ Project plans/objectives
▪ Project datasets
▪ Project codes
  ○ Jupyter notebook
  ○ R scripts
  ○ Python scripts
▪ Output files
  ○ Visualizations
  ○ Tables
  ○ Other useful outputs
▪ Project report
	▪ README file
	▪ Project datasets
	▪ Project codes
	  ○ Jupyter notebook
	  ○ R scripts
	  ○ Python scripts
	▪ Output files
	  ○ Visualizations
	  ○ Tables
	  ○ Other useful outputs
	▪ Project report

https://gist.github.com/ericmjl/27e50331f24db3e8f957d1fe7bbbe510

But why organize
your
1st Data Science Project?

- Organization increases productivity as avoid wasting time searching for project files such as datasets, codes, output files, and so on.
- A well-organized project helps you to keep and maintain a record of your ongoing and completed data science projects.
- Completed data science projects could be used for building future models.
- A well-organized project can easily be understood by other data science professionals when shared on platforms such as Github.

But

which projects to start ?

- Beginners Section [B] : Brush your basic concepts and revise them to start doing projects

Titanic Dataset

Iris Dataset

Stock Price Prediction

Stores Sales Forecasting

Housing Price Prediction

Guide for Beginner Projects:

First of all see Below 2 videos to get an idea on how to make projects of Data Science and Machine Learning And then Move to Kaggle for Making your own project.Its is Good if you Make Minimum 2-3 Projects on your own.

- Titanic Survivor : https://www.youtube.com/watch?v=fS70iptz-XU&t=
- Credit Card Fraud Detection : https://www.youtube.com/watch?v=gCWBFyFTxVU

Intermediate & Advanced Section

- Learn libraries like Opencv , Tensorflow , SkLearn

1) Natural Language Processing : MNIST Handwritten Digit Classification , Twitter Sentiment Analysis

2) Email Spam Classifier

3) Fraud Detection System

4) Computer Vision : Face Recognition , Face Detection

Ayon Roy

Speaker 🏆 Let's talk ML, AI, Data Science, DL, Python 🏆 Catch me @ Hackat...

1w •

3 Major types of projects you should do if you are just diving into **#datascience**, **#machinelearning**, **#artificialintelligence**. Here are a few pointers :

For Exploratory Data Analysis (EDA) Projects -
Practice on the dataset at
- **https://lnkd.in/gztCfy3**
- **https://lnkd.in/gFasqNi**
- **https://lnkd.in/grvF-jc**
- **https://lnkd.in/gPsxf5y**
- **https://lnkd.in/gDKuhEf**
- **https://lnkd.in/g_SRS7F**

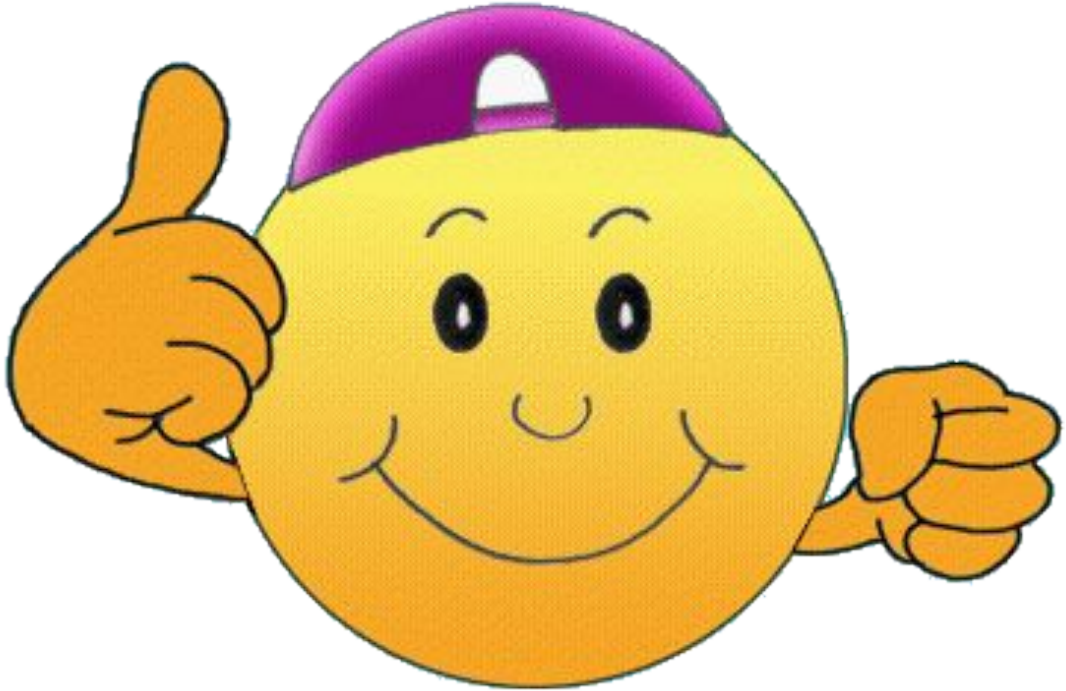
For Prediction Modelling Projects -
Practice on the dataset at
- **https://lnkd.in/gQh6SRZ**
- **https://lnkd.in/g5JfbeA**
- **https://lnkd.in/gPG6Wgf**
- **https://lnkd.in/gYBE6DY**

For Data Visualization Projects -
Practice on the dataset at
- **https://lnkd.in/gWZJ3TZ**
- **https://lnkd.in/gih7YDd**
- **https://lnkd.in/gcv2xar**

A few useful resources

- https://towardsdatascience.com/how-to-plan-and-organize-a-data-science-analytics-project-a9418c12c808
- https://towardsdatascience.com/how-to-organize-your-data-science-project-dd6599cf000a
- https://ayonroy.ml/help

GO FOR IT !



GOOD LUCK !

Let me answer your Questions now.

Finally, it’s your time to speak.

Danke Schoen

Questions ? Any Feedbacks ? Did you like the talk?
Tell me about it.

If you think I can help you,
connect with me via

Email : ayon.roy2000@gmail.com

LinkedIn / Github / Telegram Username : ayonroy2000

Website : https://AYONROY.ML/