

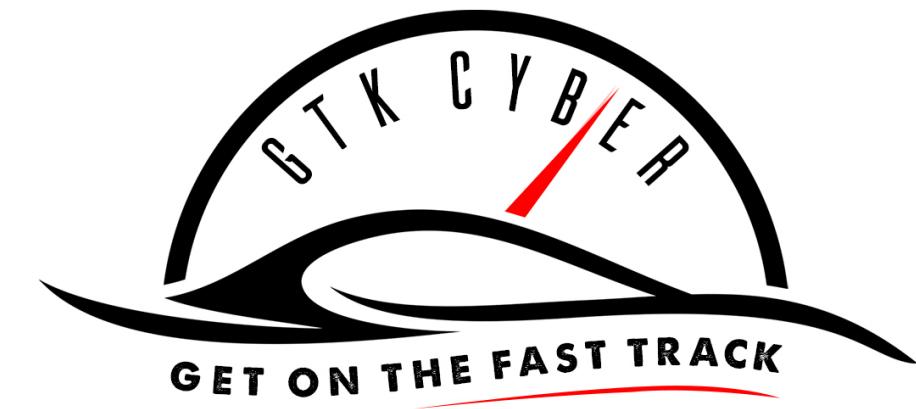
GTK CYBER

GET ON THE FAST TRACK

Applied Data Science for Cyber Security Professionals

www.gtkcyber.com

Join the Slack Channel at: <http://bit.ly/30IvONs>



Course Agenda

Day 1: Introduction & Data Engineering

- Intro: What is Data Science?
 - Overview of Machine Learning & Cyber Applications
- Regular Expressions
- Manipulating and Exploring Data
 - Exploratory Data Analysis in 1 Dimension

Day 2: Data Engineering & Data Exploration

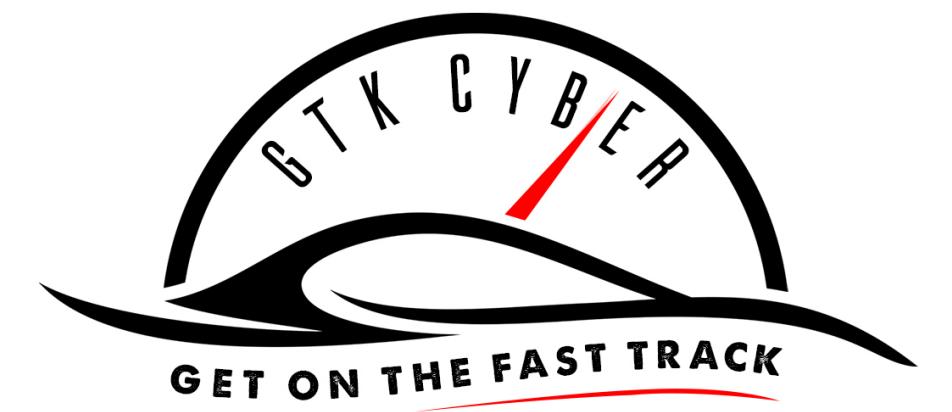
- Exploratory Data Analysis in 2 Dimensions
- Rapid Data Exploration with Apache Drill
- Data Visualization

Day 3: Machine Learning

- Feature Engineering
- Supervised Machine Learning
- Unsupervised Machine Learning
- Hacking Machine Learning Models

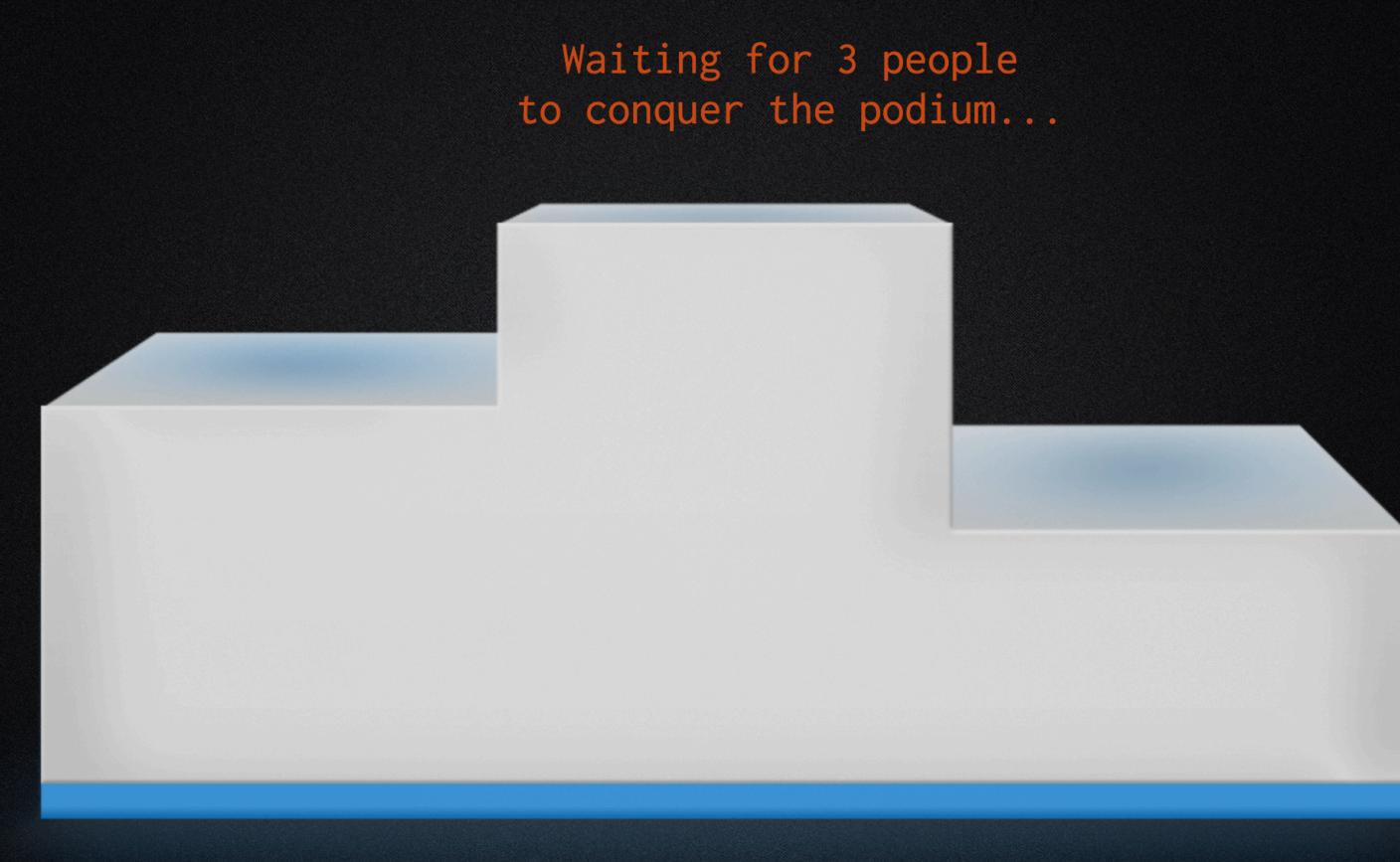
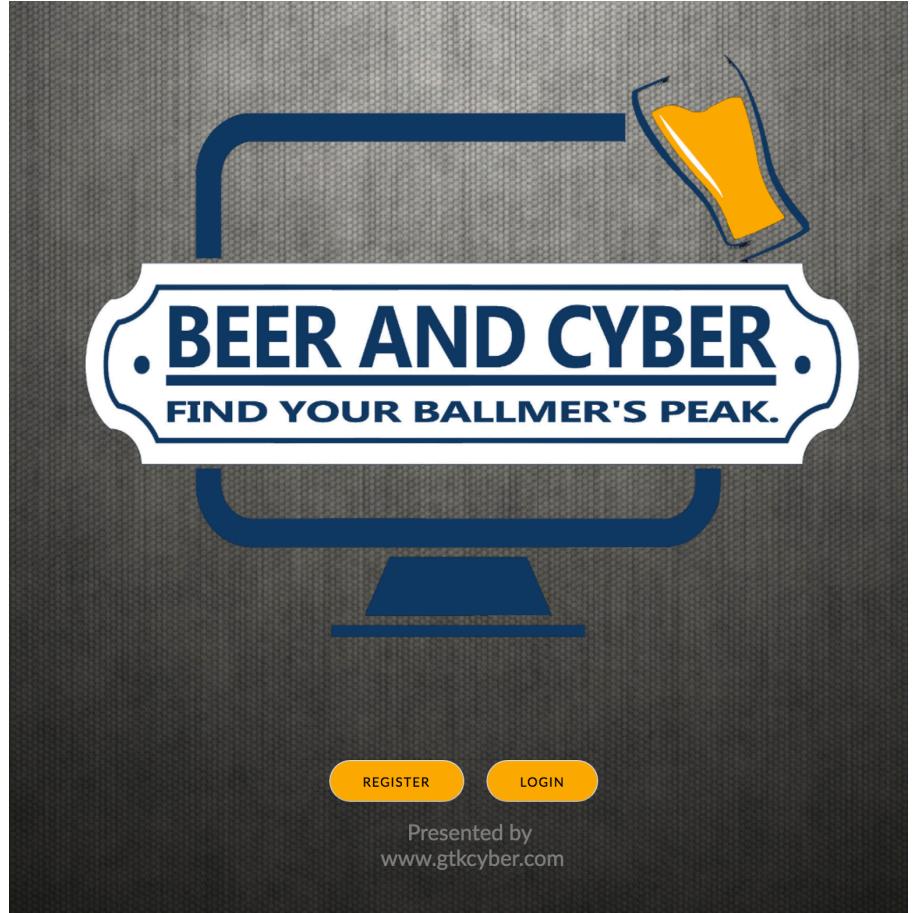
Day 4: Advanced Topics

- Introduction to Big Data Tools
- Anomaly Detection
 - PySpark, ELK & Kafka
- Hunting with Data Science



Beer & Cyber CTF

- www.beerandcyber.com
- username: blackhat
- password: 2019
- Register and start!



Challenges

- ▲ Warm Up
- ▲ Easy
- ▼ Medium
 - Datetimes-1 (50 pts)
 - Encoding/Counting (50 pts)
 - Data selection (50 pts)
 - String manipulation (50 pts)
 - HTTP APIs, JSON, XML (50 pts)
 - SHA512 Hashes (50 pts)
 - Parsing Logs 1 (50 pts)
 - Parsing Logs 2 (50 pts)
- ▲ Hard
 - 🔒 H4ck3r

Parsing Logs 1

Question
Given two sets of **apache log files** find the most popular IP. Separate IPs with a comma.

Example: 192.168.0.1, 192.168.0.10

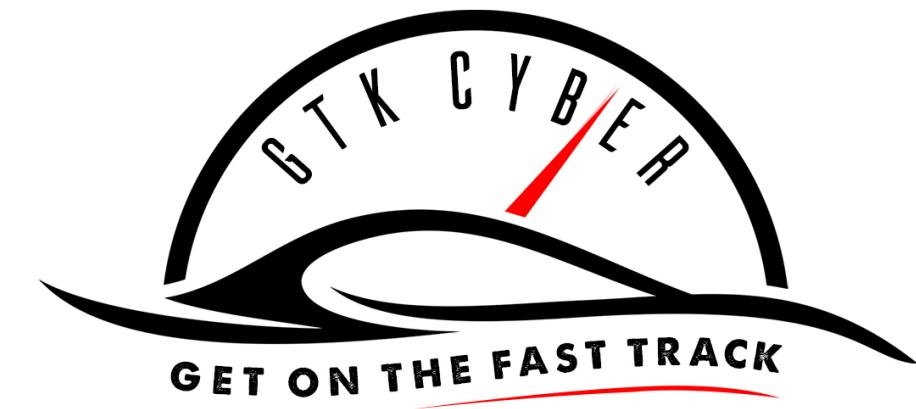
Submit Flag (50 pts)

Submit

Hint Available

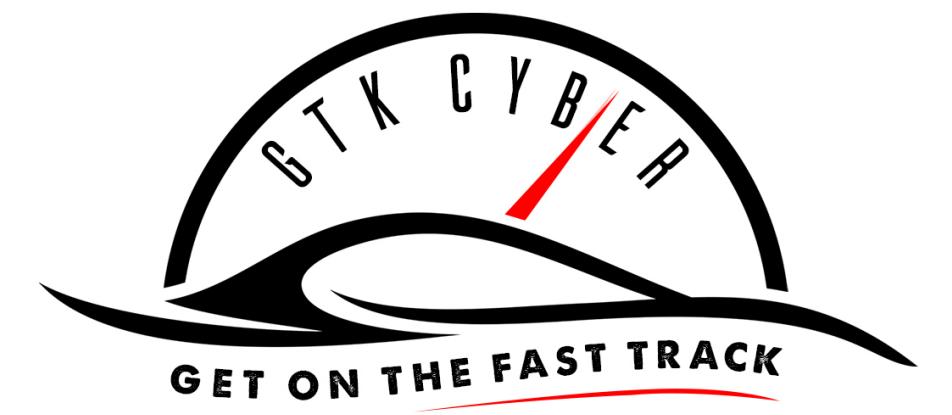
Previous Submissions

Interactive Shell

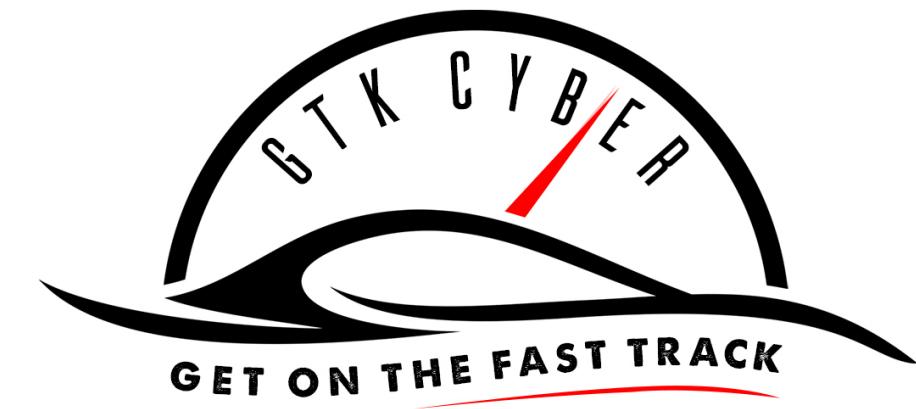


Expectations

- Please participate and **ask questions**.
- Please follow along and **TRY OUT** the examples yourself during the class
- All the answers are in the slide decks or GitHub repository, but please try to complete the exercises **without looking at the answers**.
- Join the conversation in slack!
- Have fun!



Introduction

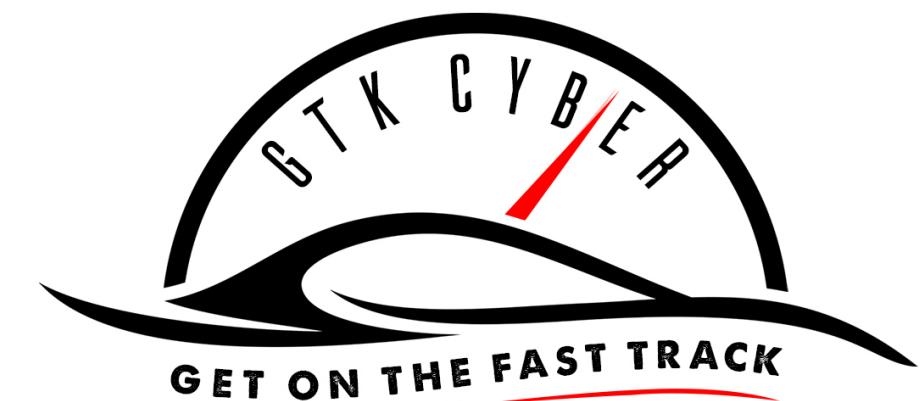


Our Lawyers Make Us Say This



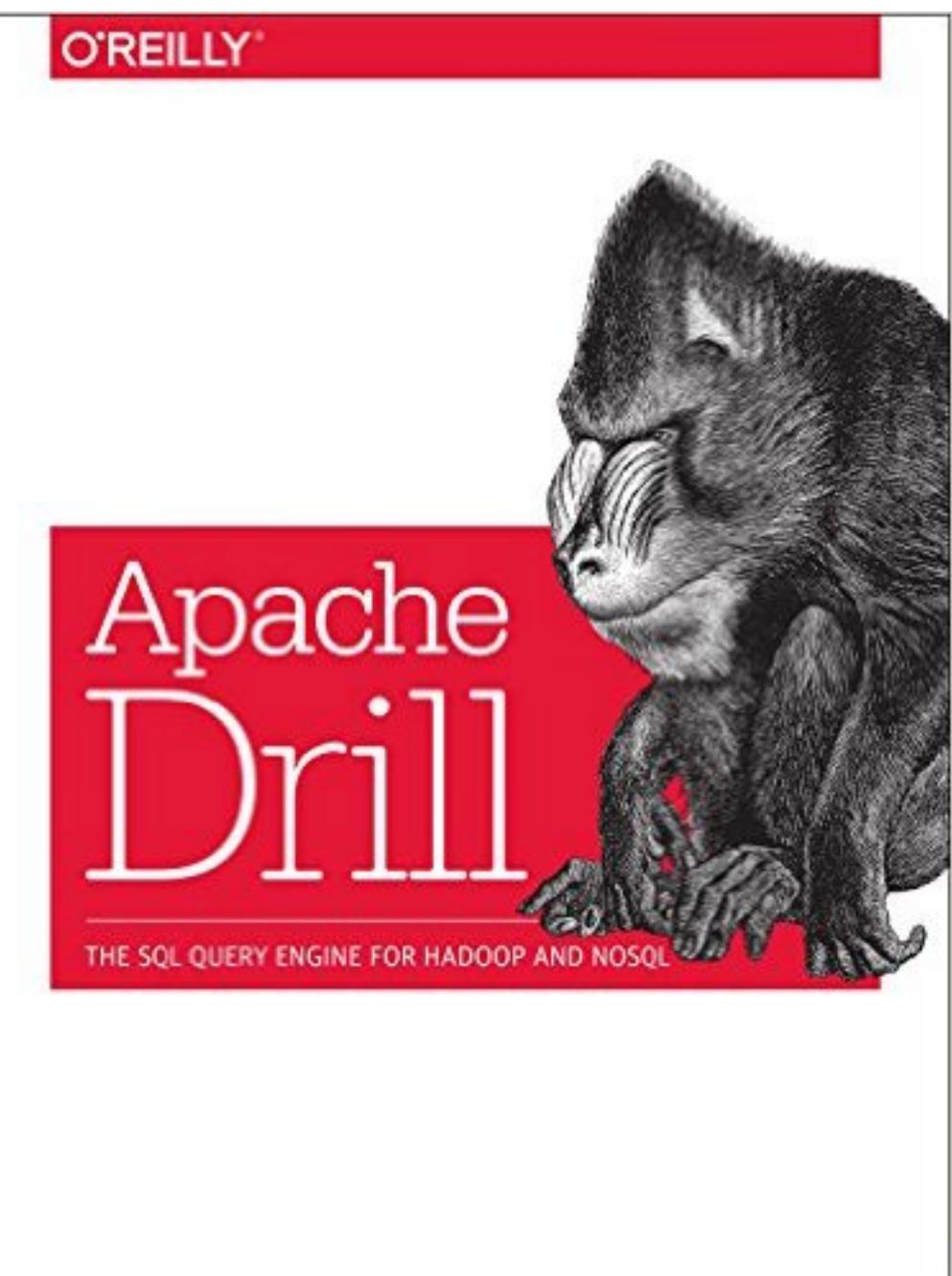
All materials presented in this training and those provided as an adjunct to the program are copyrighted 2019 by GTK Cyber LLC.

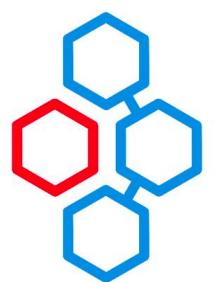
They are intended solely for the use of registered program participants and may not be reproduced or redistributed in any manner for any other reason.



Charles Givre, CISSP

- Lead Data Scientist at Deutsche Bank
- PMC Chair for Apache Drill
- Senior Lead Data Scientist @ Booz Allen
- 5 Years @ CIA
- Undergraduate in Comp.Sci & Music





Austin Taylor

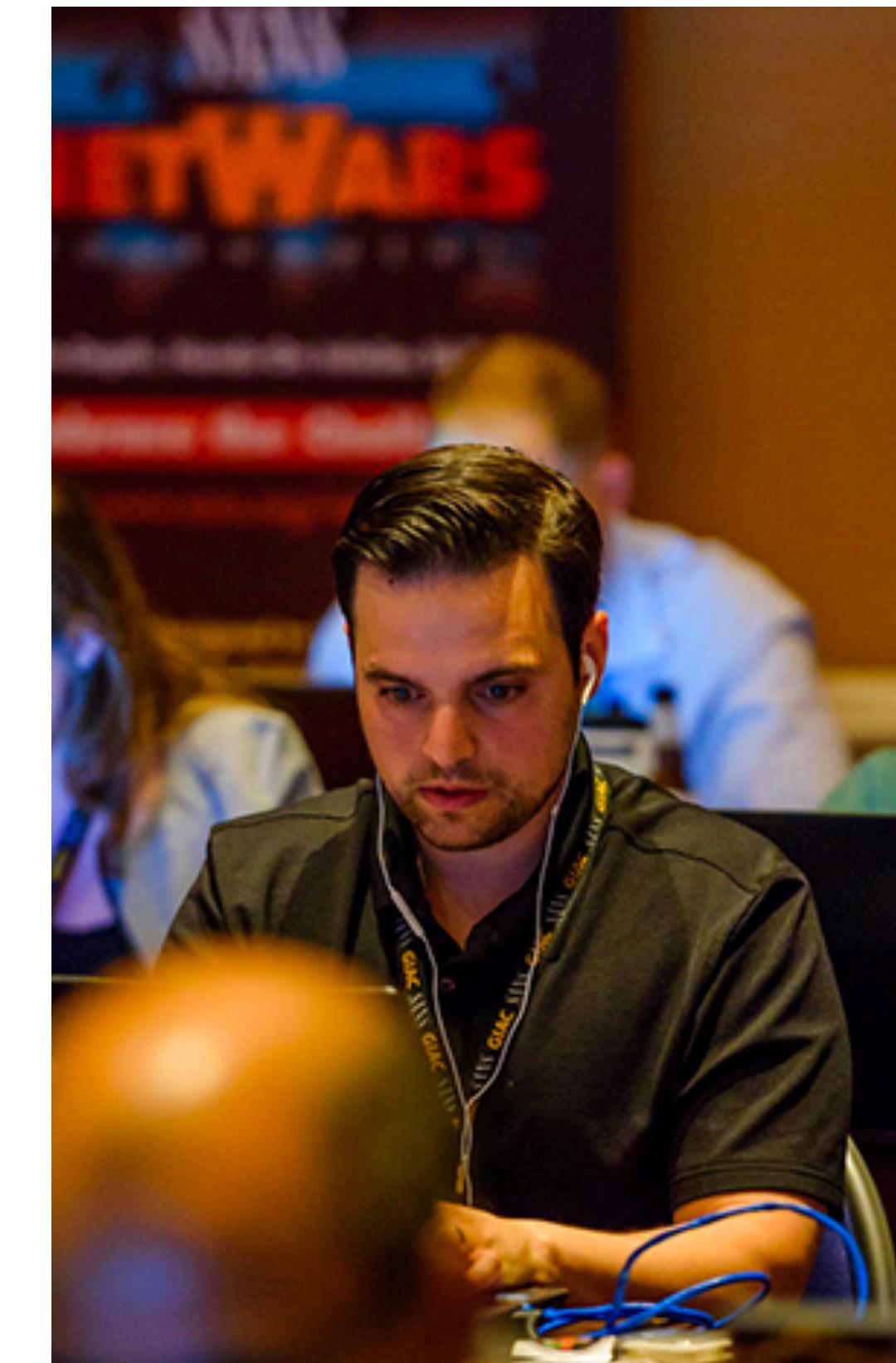


Director of Cybersecurity R&D
@IronNetCyber

Cyber Warfare Operations Officer
@USAF

<https://www.github.com/austin-taylor>

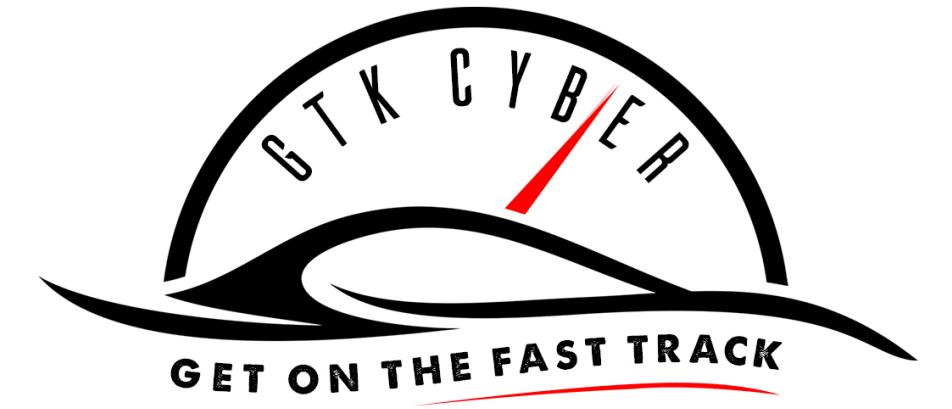
- VulnWhisperer
- Flare
- Bluewall



www.austintaylor.io

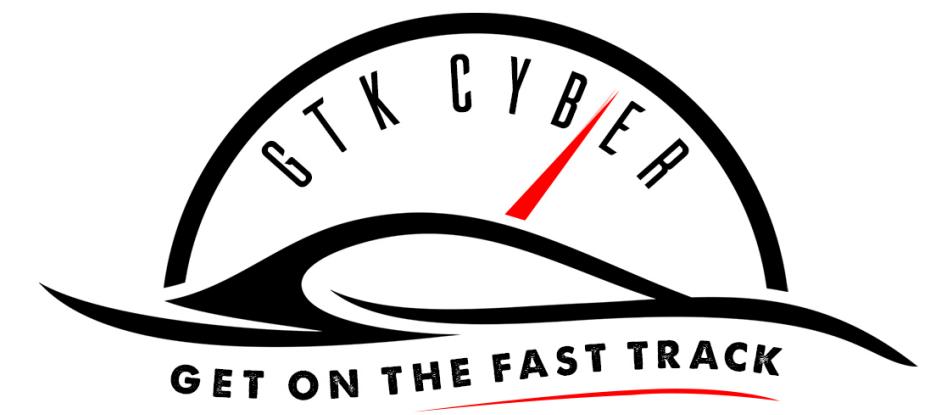


@HuntOperator

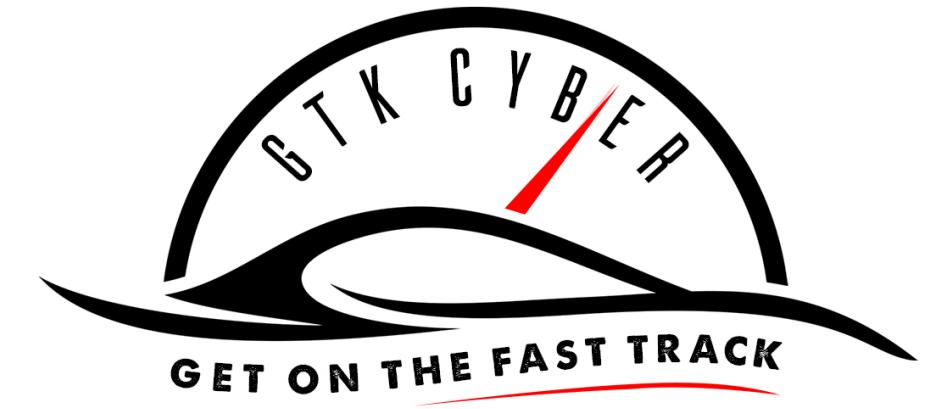


Who are you?

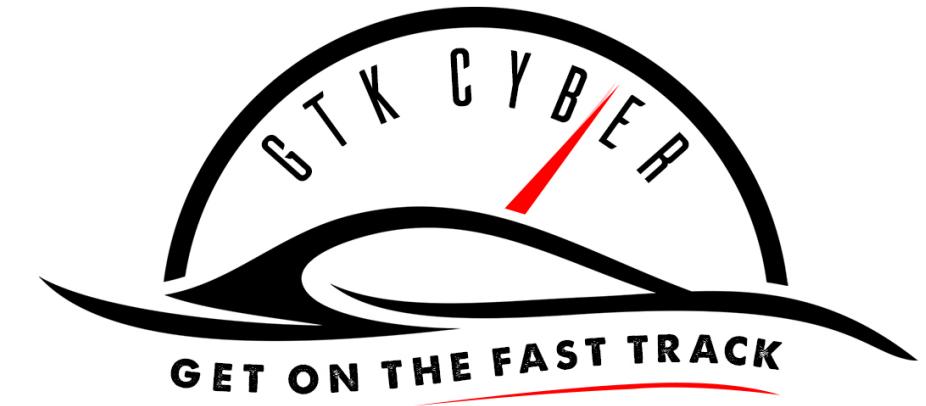
- Your name (or what you want me to call you)
- Where you work and/or your job role
- What you hope to get out of this class
- Your level of experience with coding



What is Data Science?

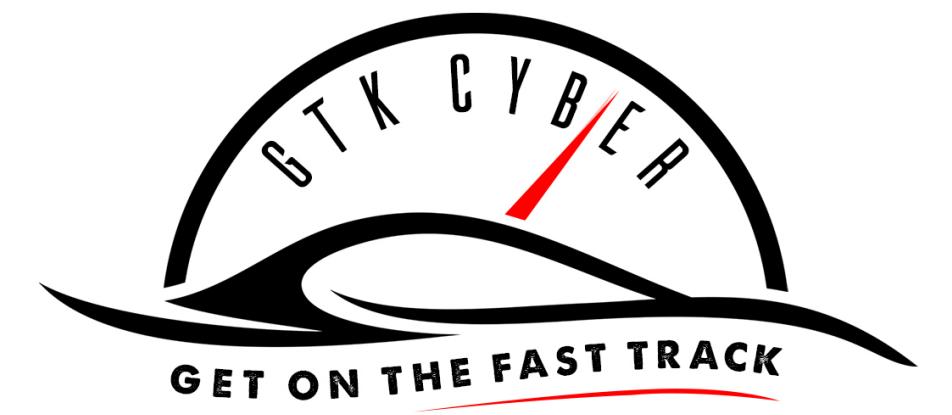


Data Science is the automated extraction of **information** from **raw data**.

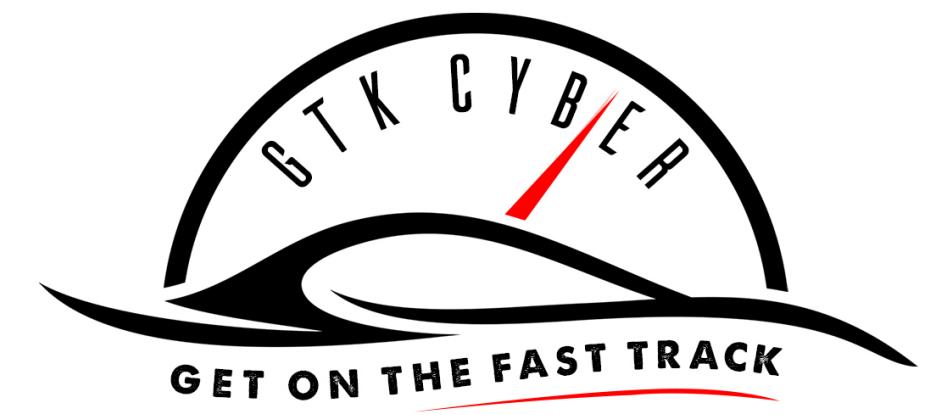


Data Science is the art of **turning data into actions**. This is accomplished through the creation of data products, which provide actionable information without exposing decision makers to the underlying data or analytics

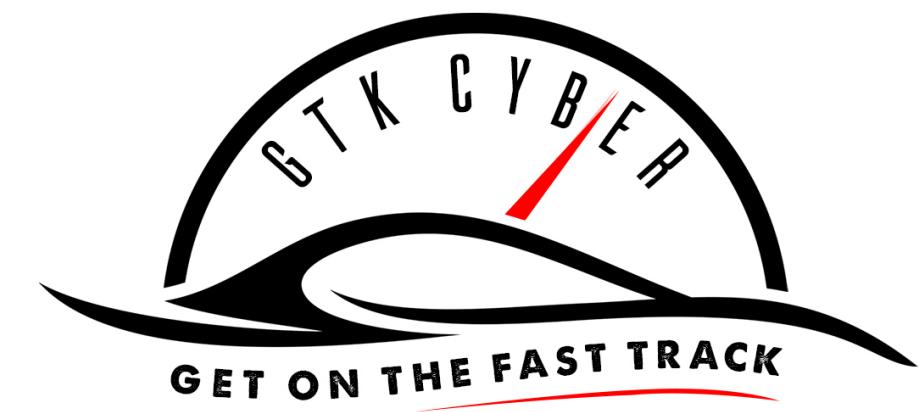
Booz Allen Hamilton, Field Guide to Data Science, Pg. 17



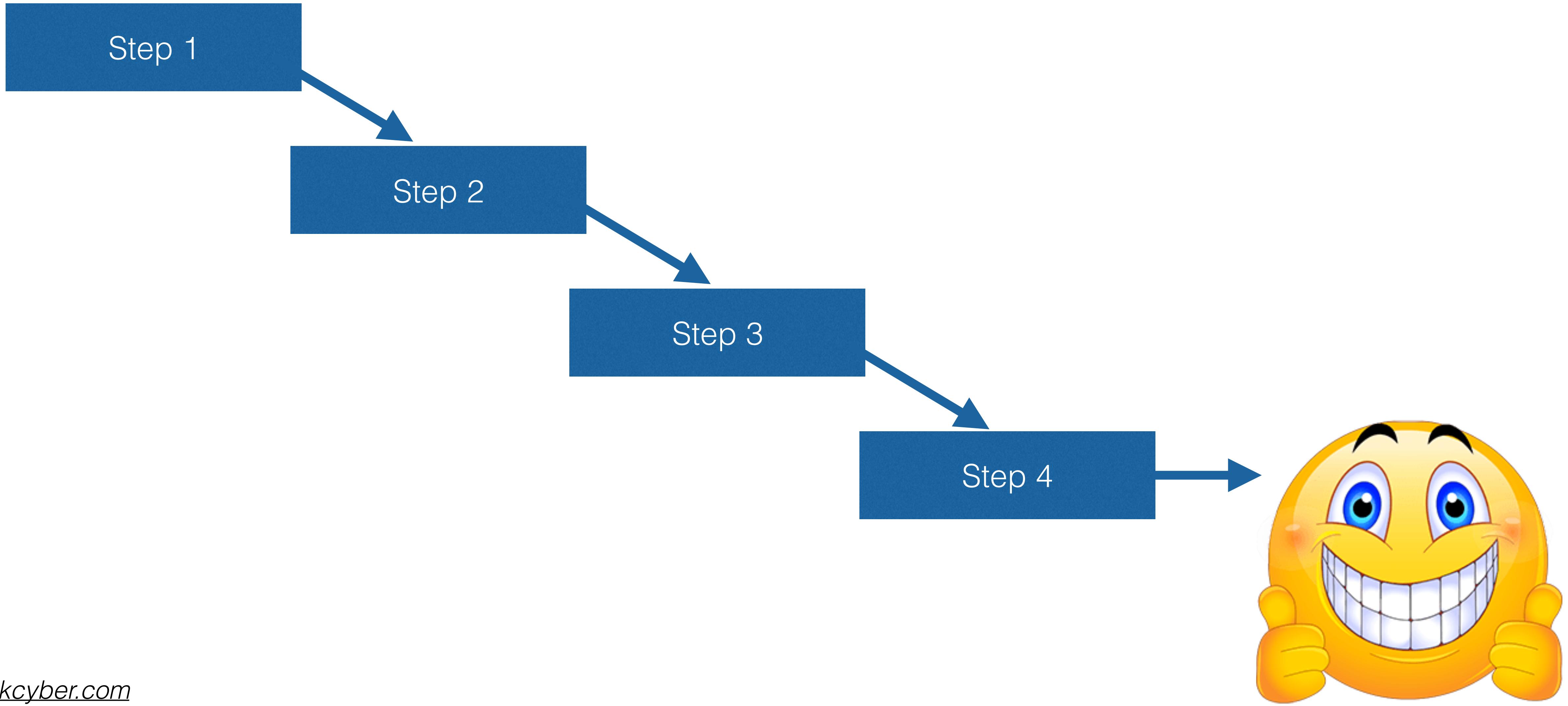
Analyst ← → Developer

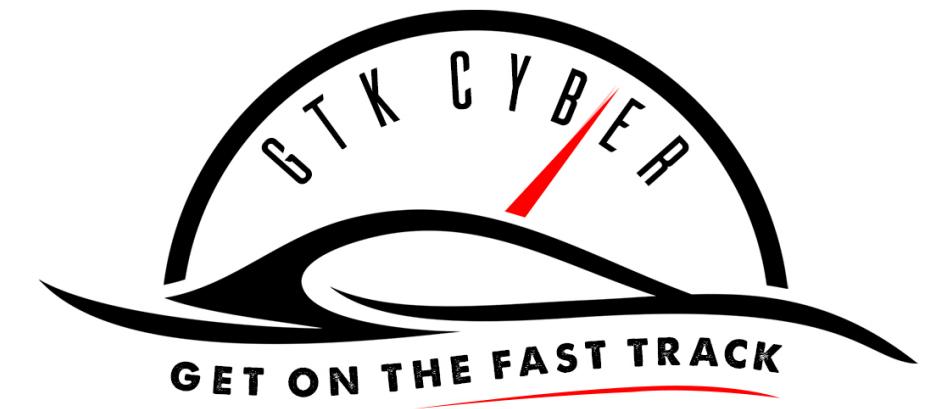


Analyst + Developer

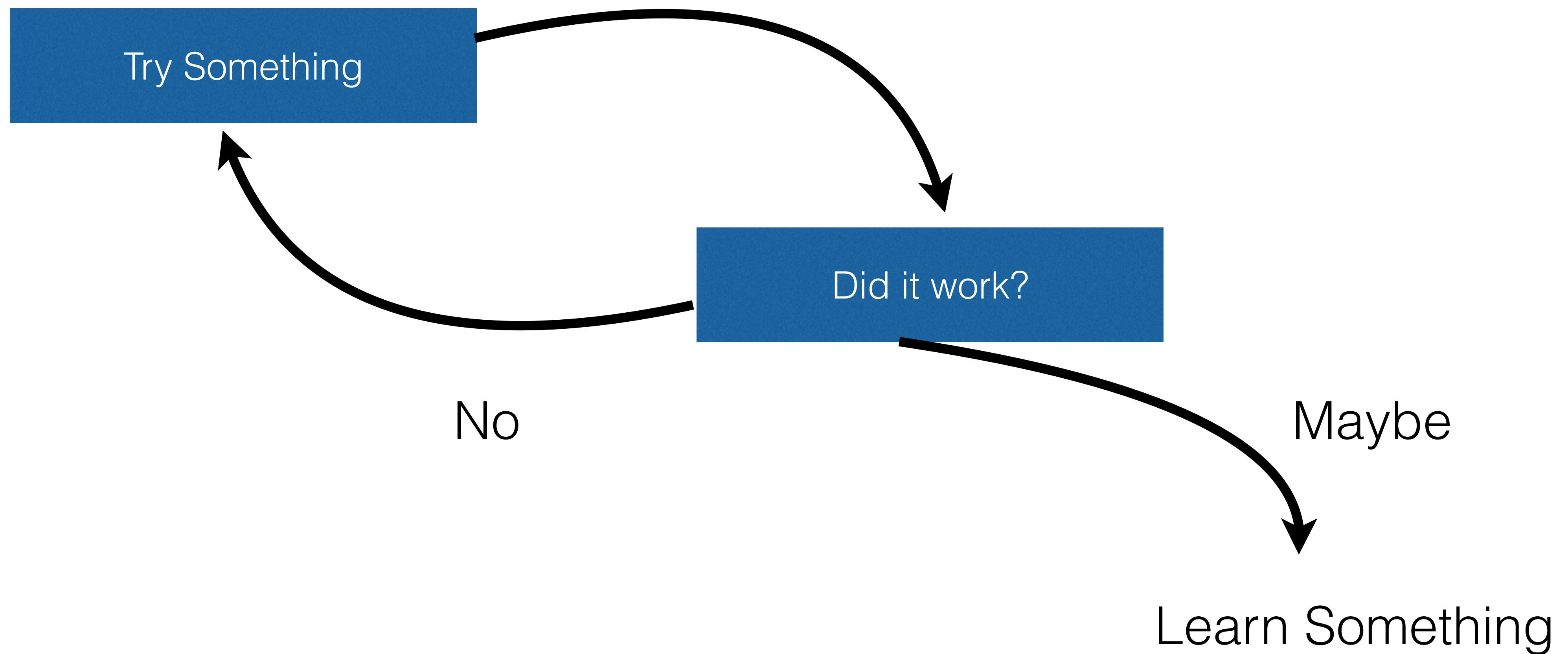


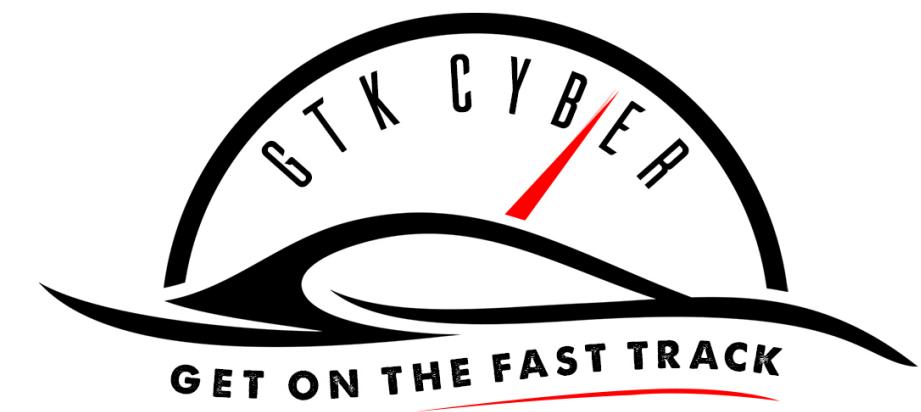
What Data Science is Not



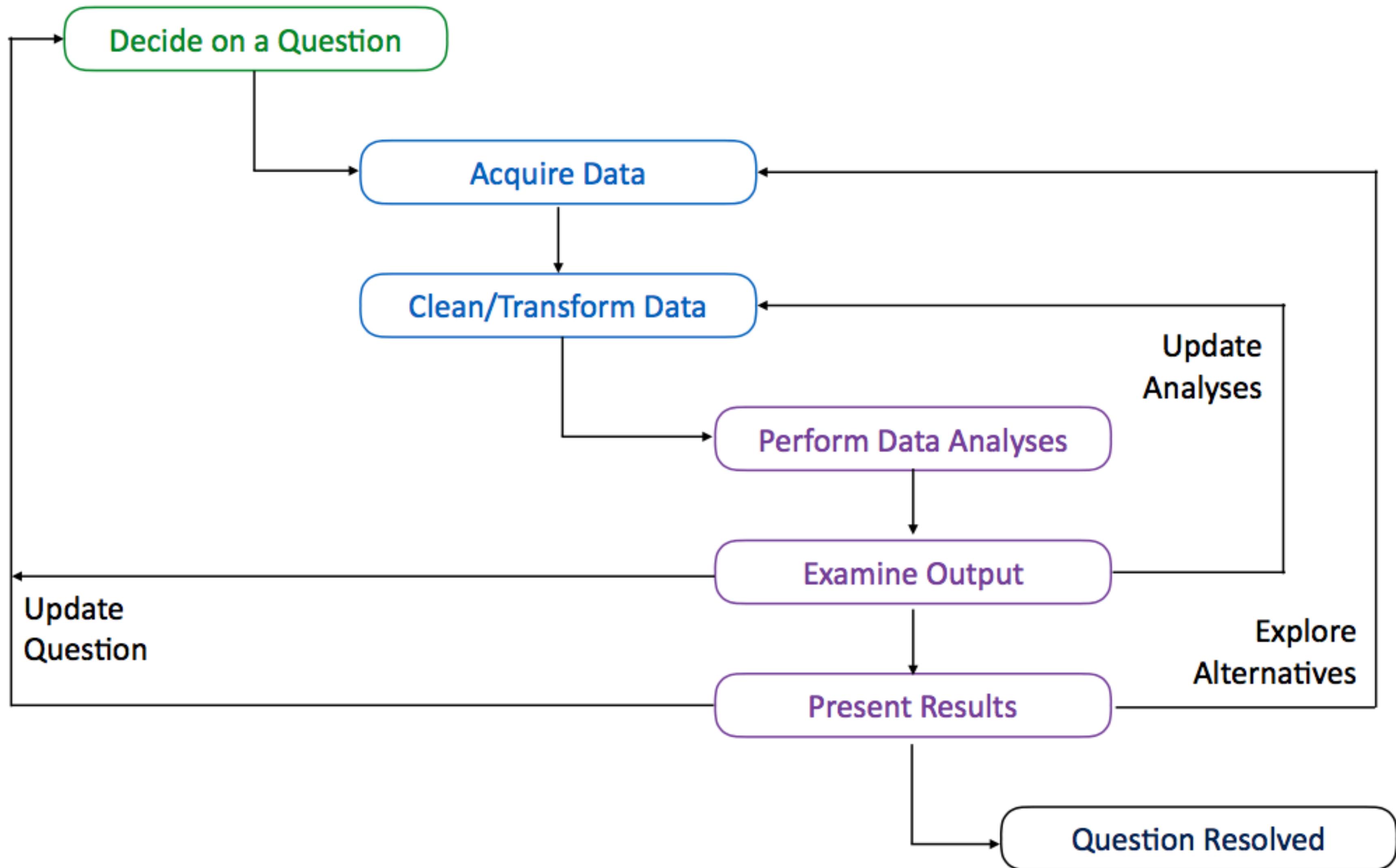


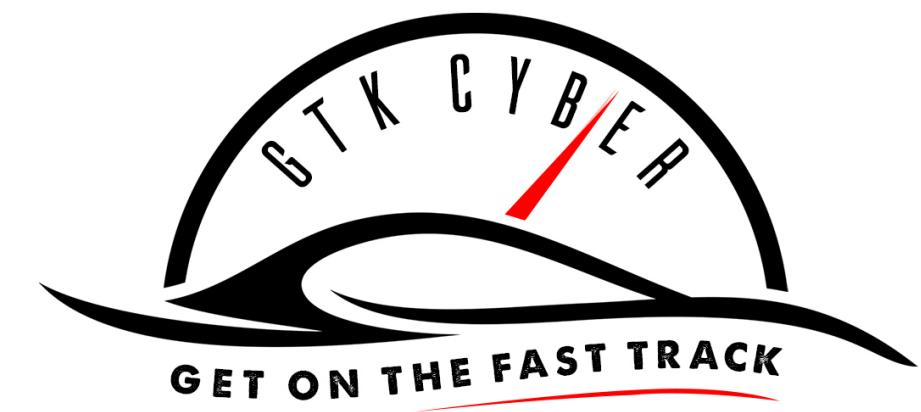
What Data Science is





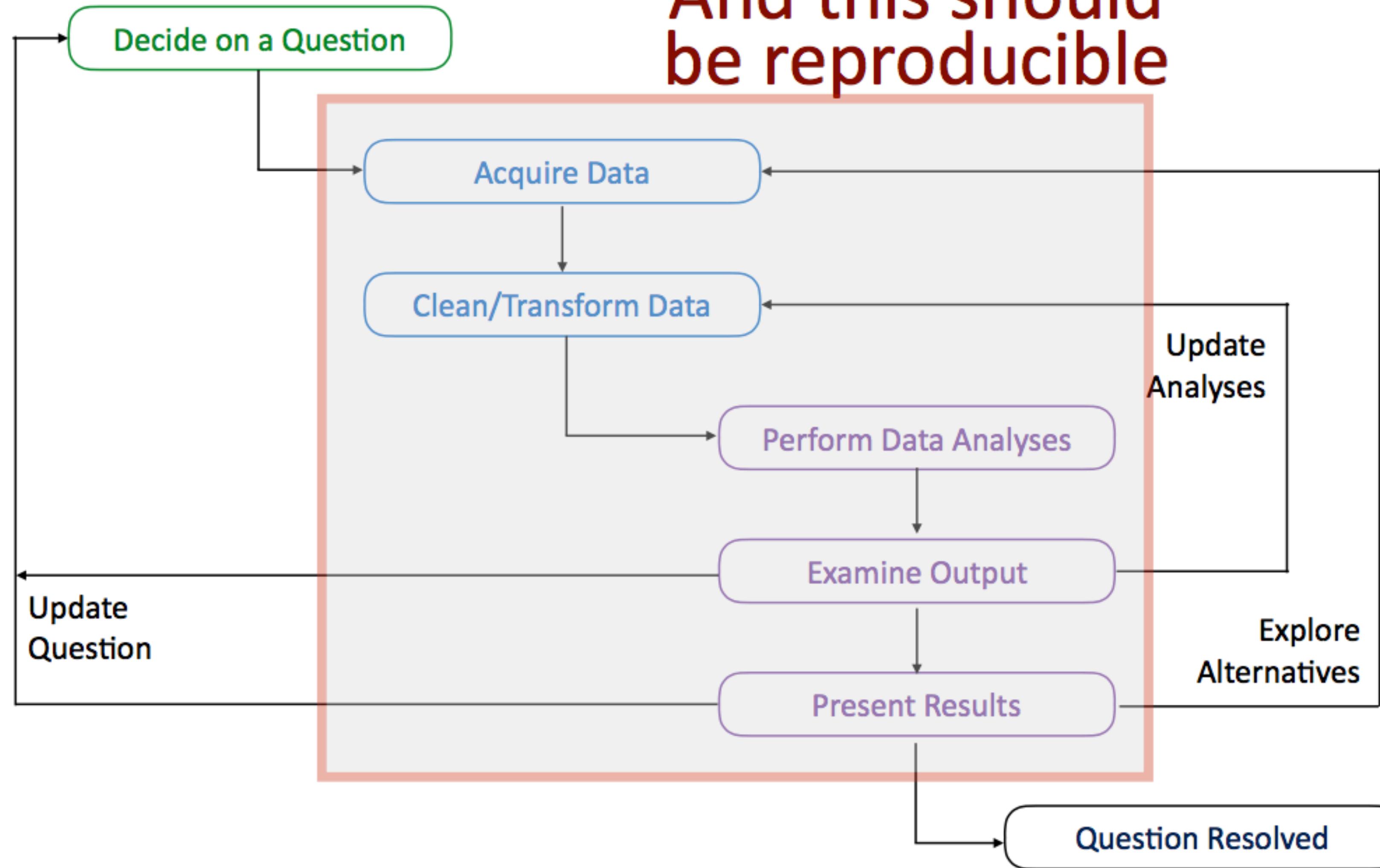
Research Process





Research Process

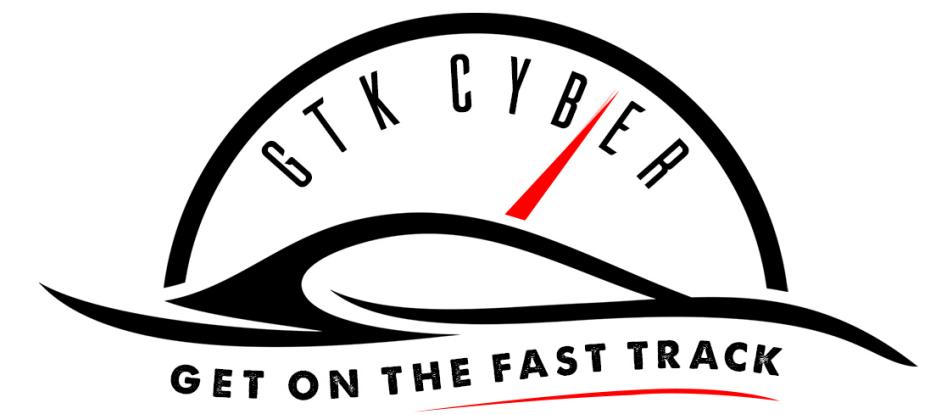
And this should
be reproducible



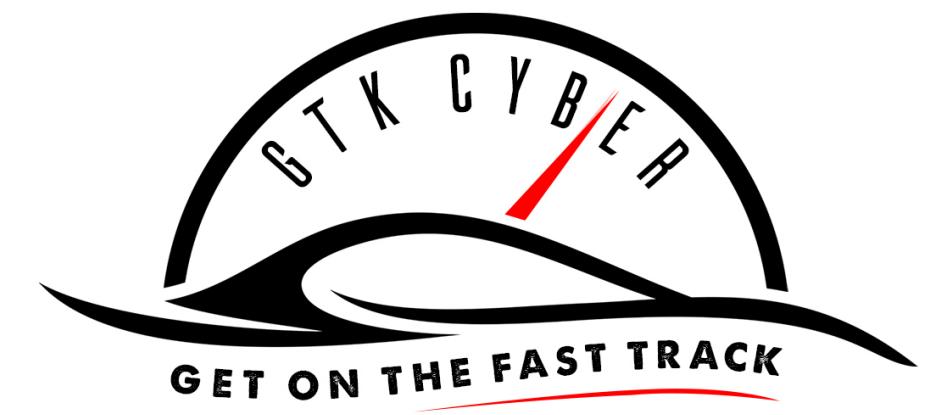


“The term "data scientist" will subside and may well sound dated five years from now. **The skills will become more commonplace and commoditized. When that happens, the real boom will begin**, because the technology will become widely adopted and thus more useful. ... **Instead, we need self-service tools that empower smart and tenacious business people to perform Big Data analysis themselves.**

–Andrew Brust, “Data scientists don't scale”, <http://www.zdnet.com/article/data-scientists-dont-scale/>

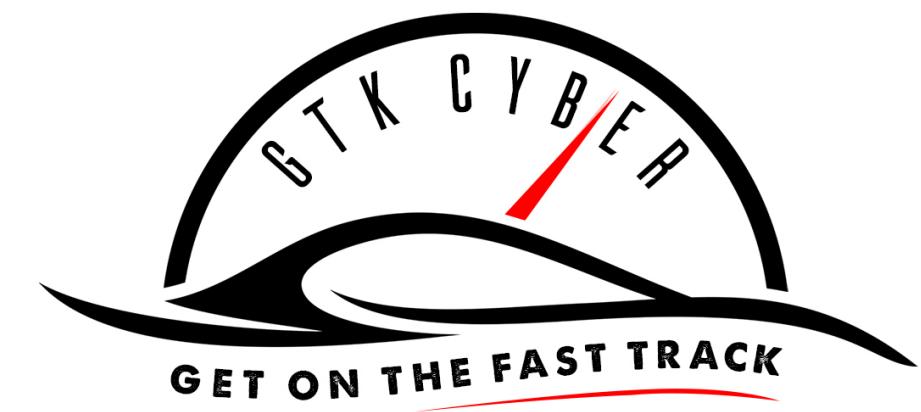


Time to Insight



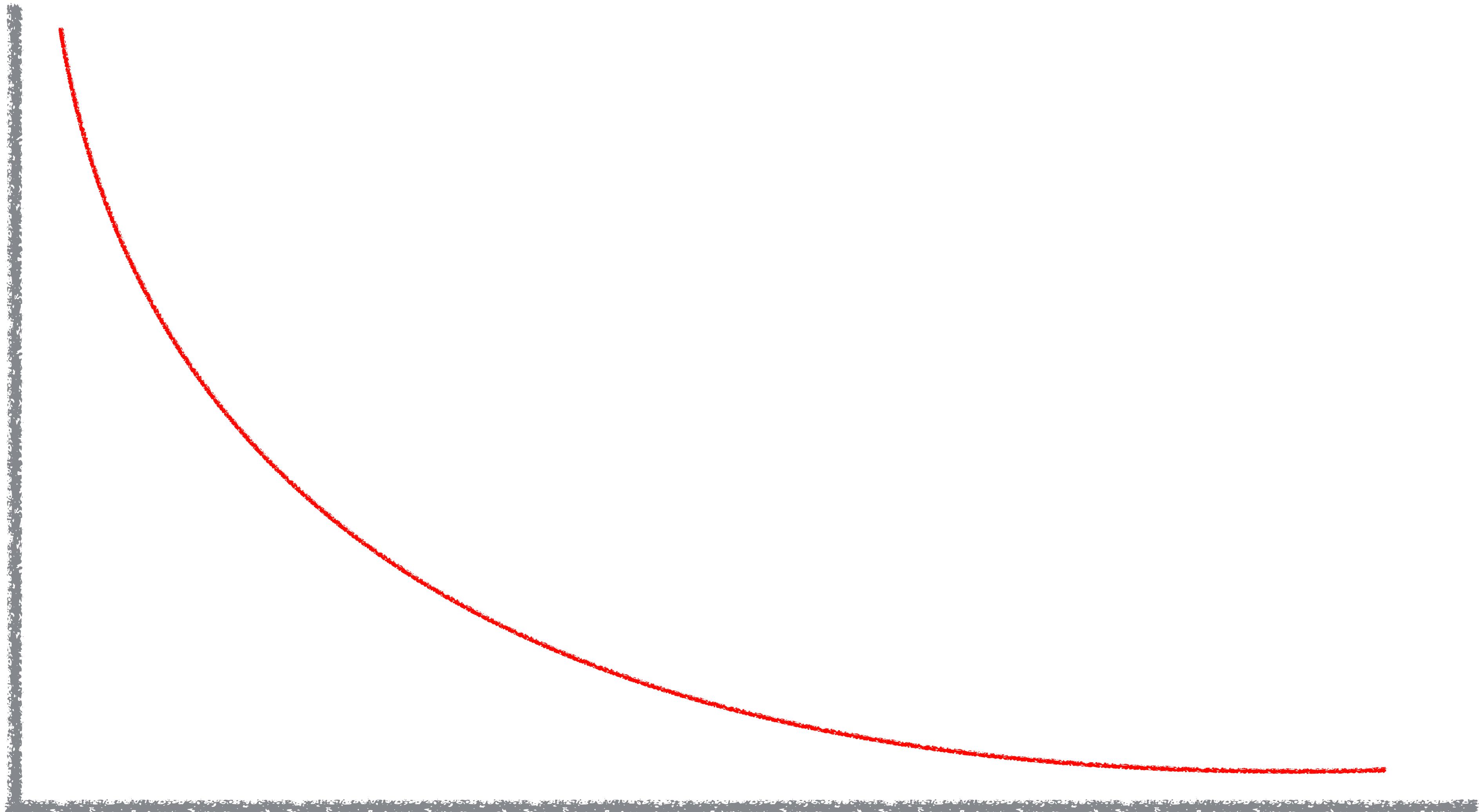
Time to Insight

Time = \$\$

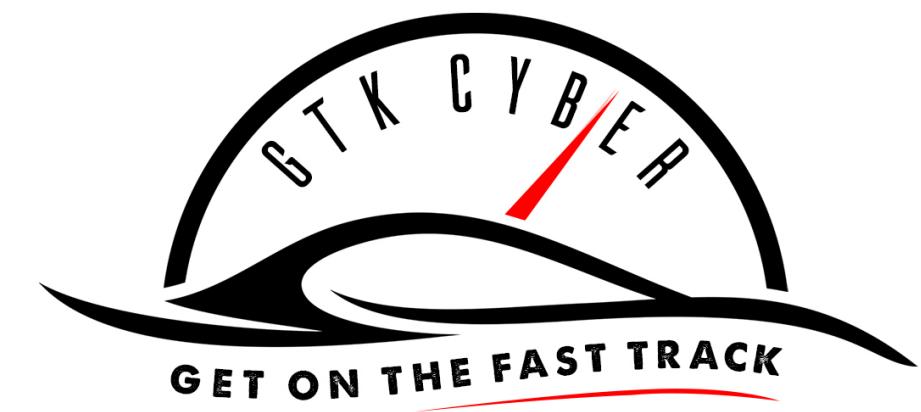


Value of Insights over Time

Value



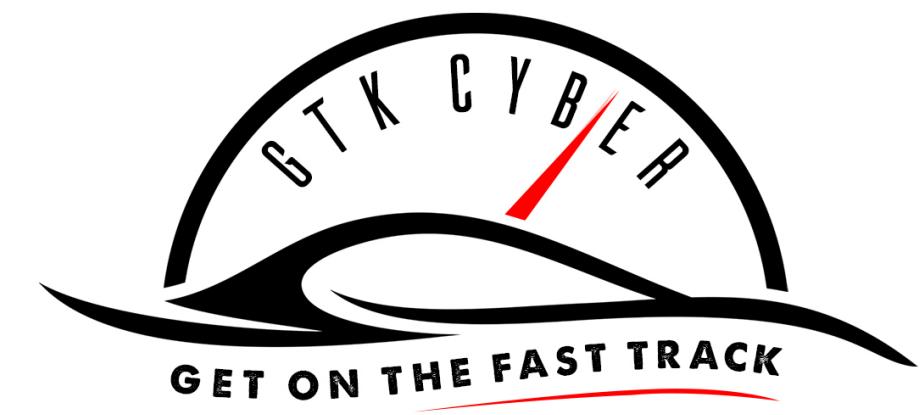
Time



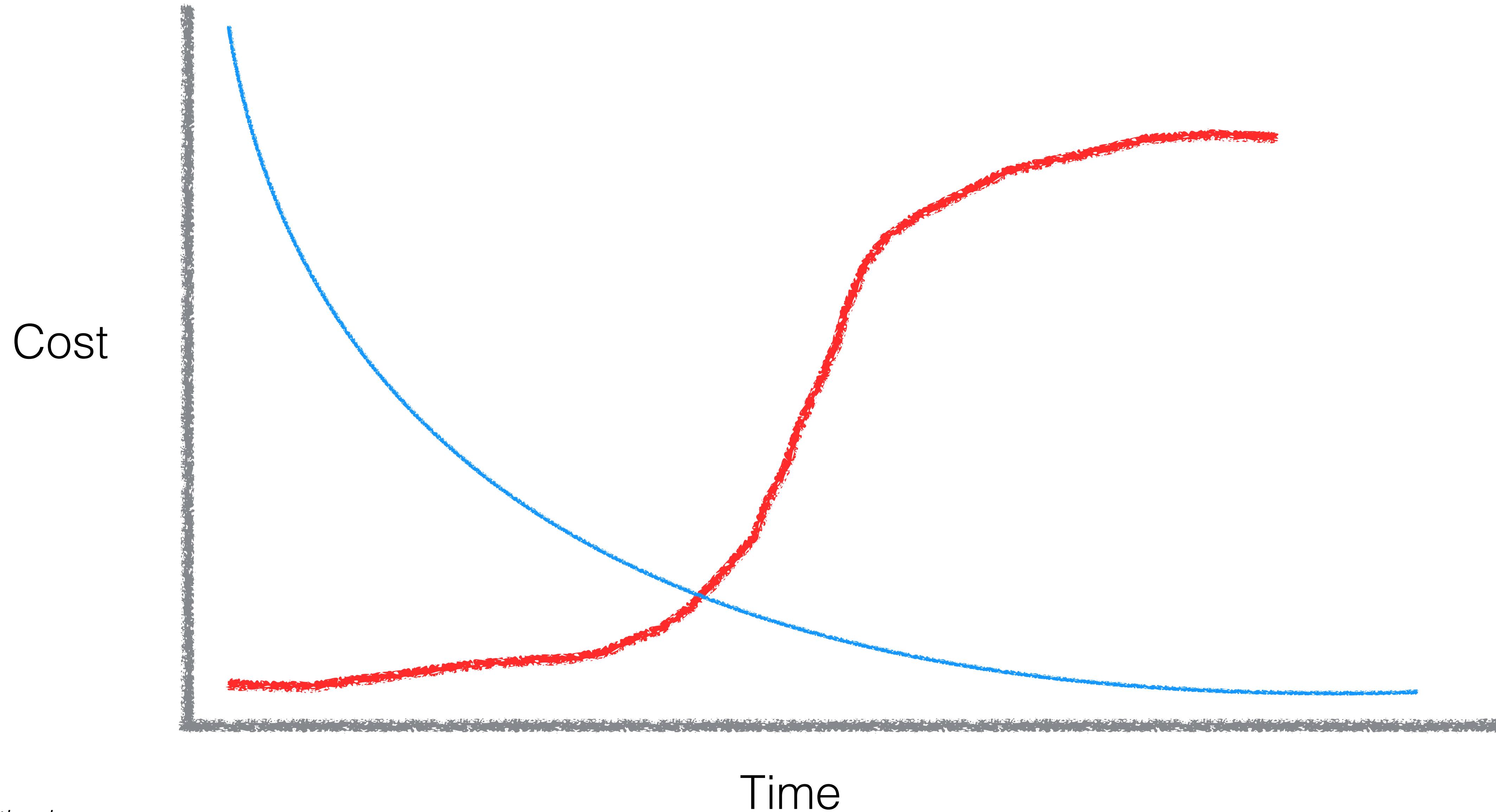
Cost of Insights over Time

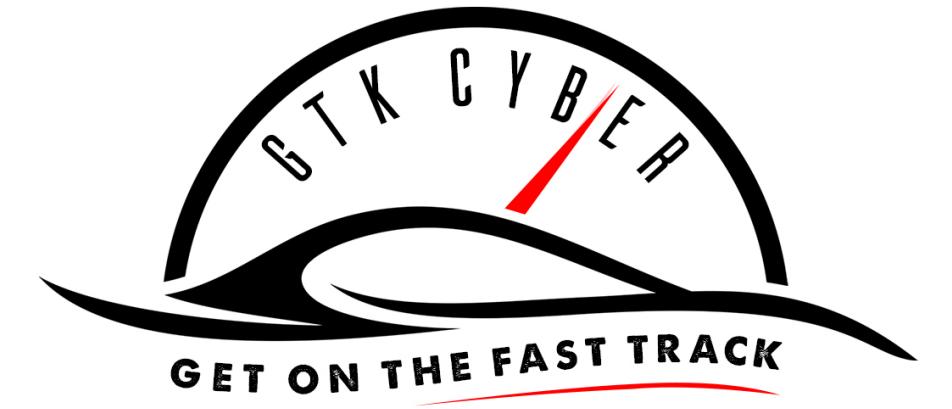
Cost

Time

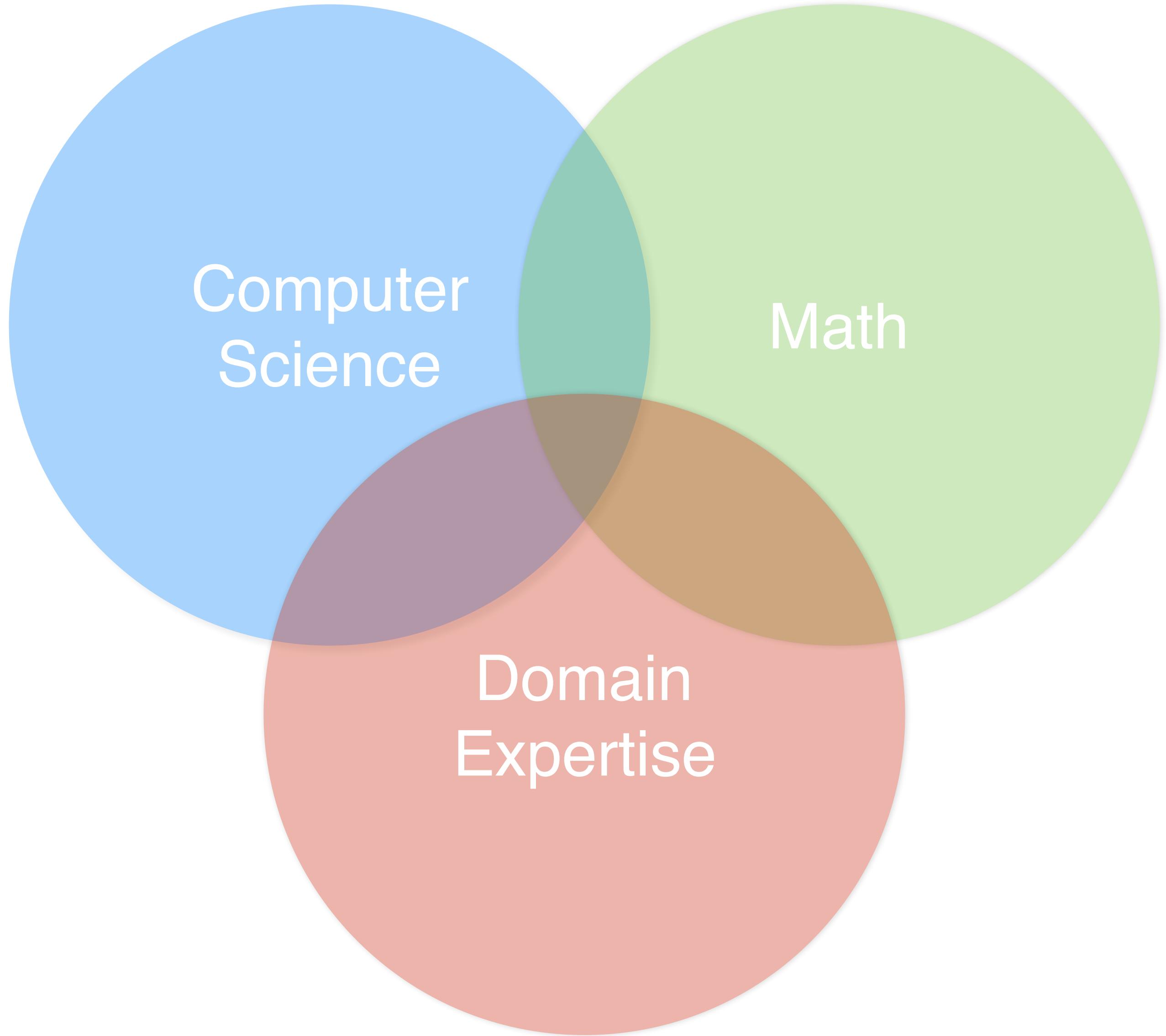
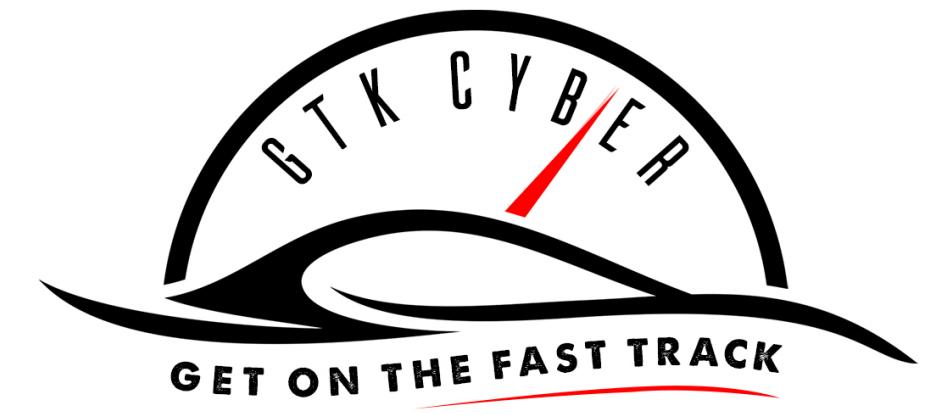


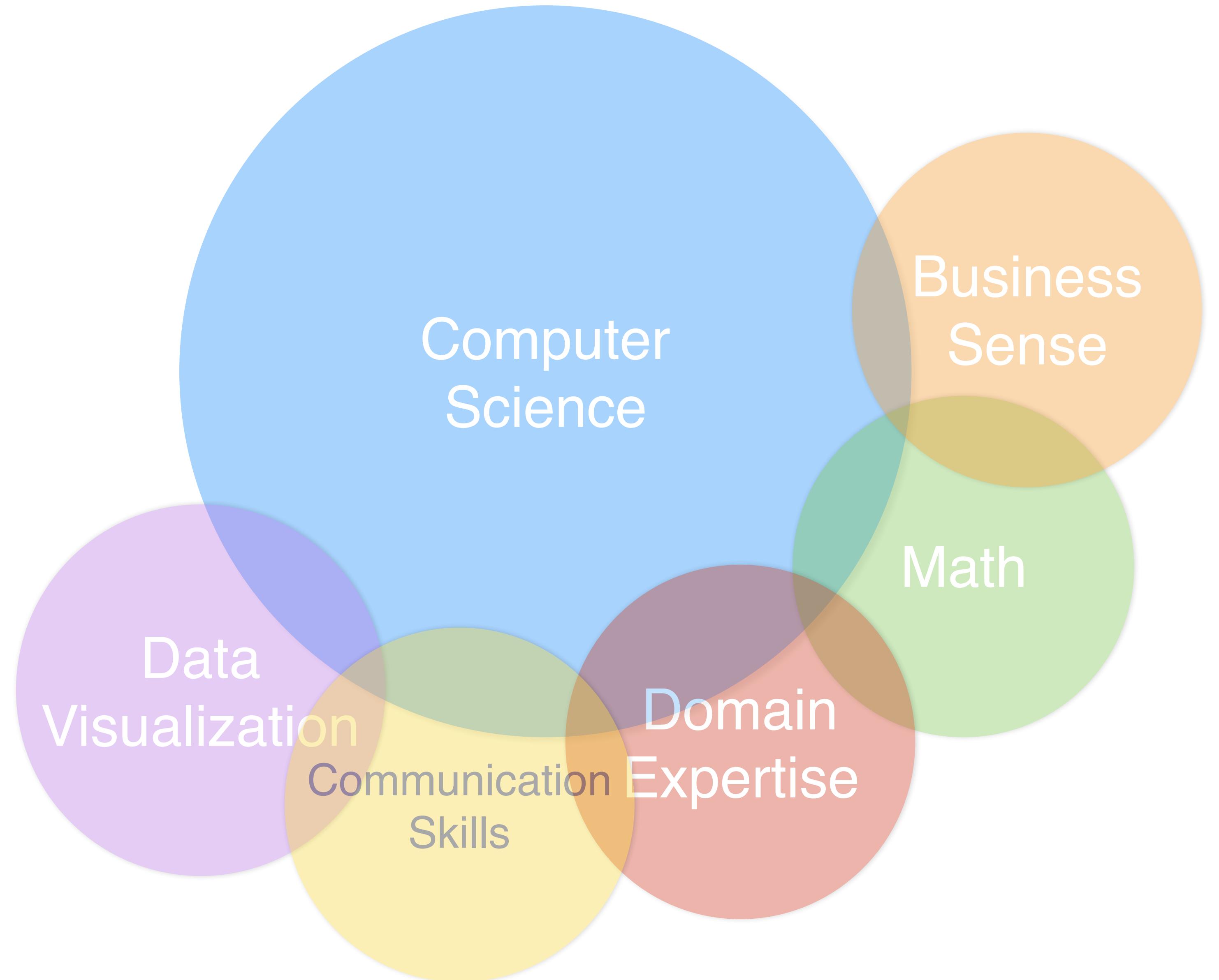
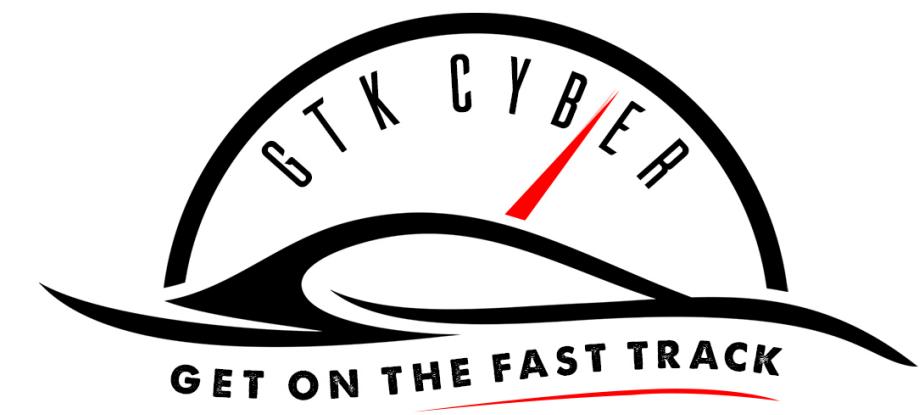
Cost of Insights over Time

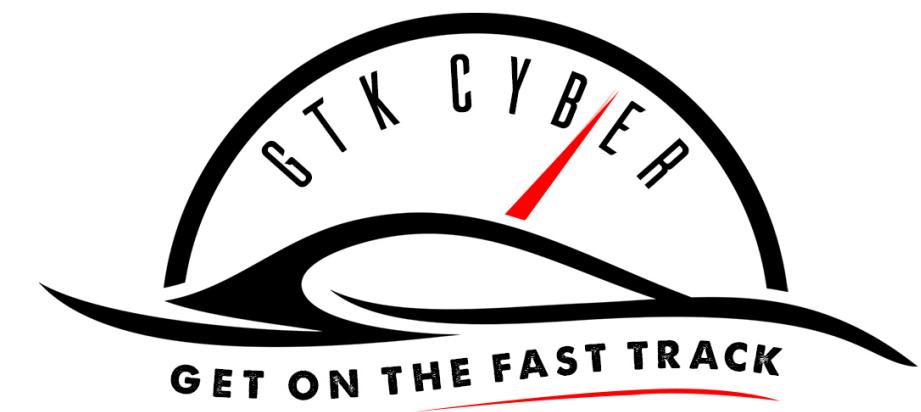




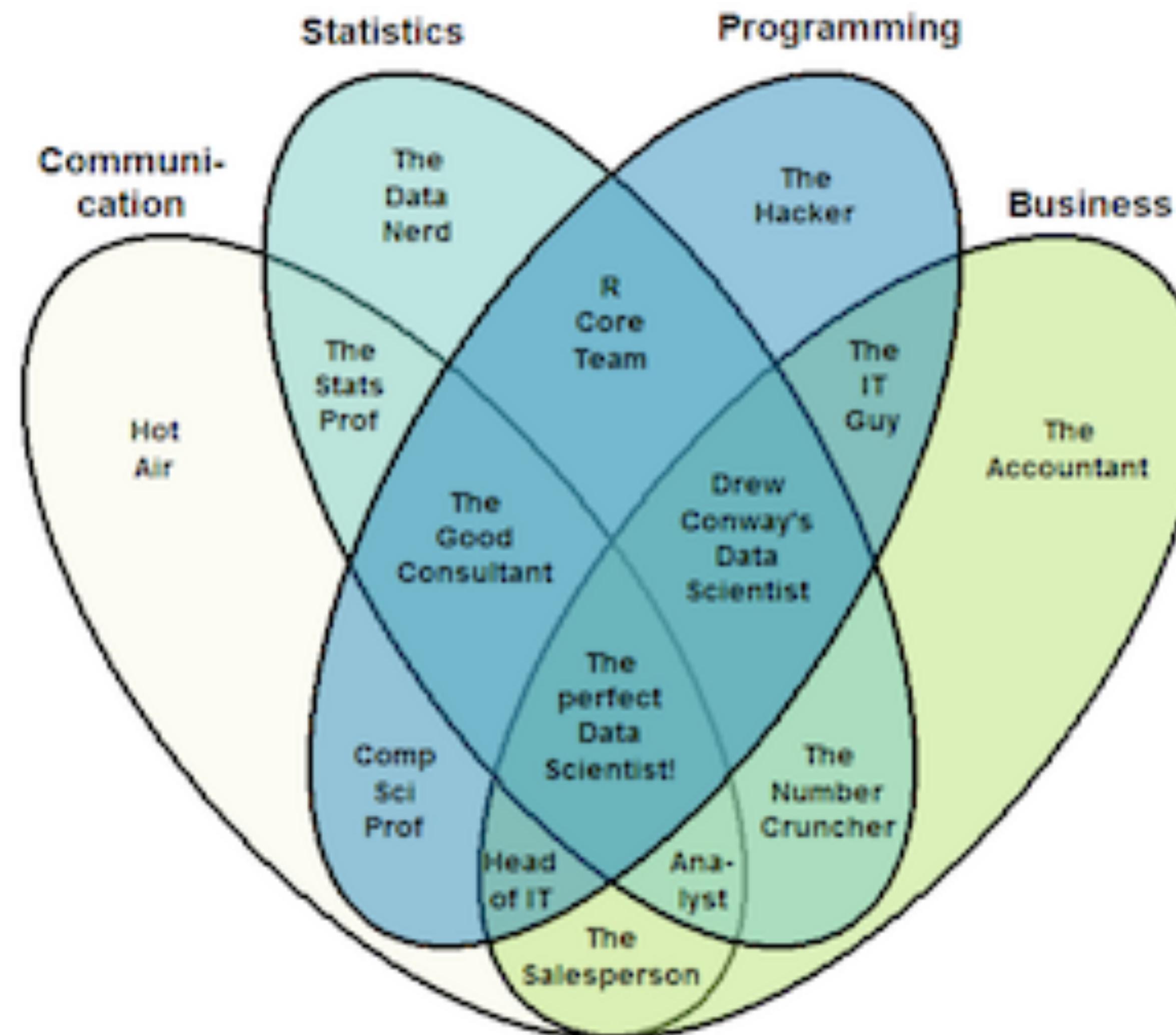
What skills does a data scientist need?

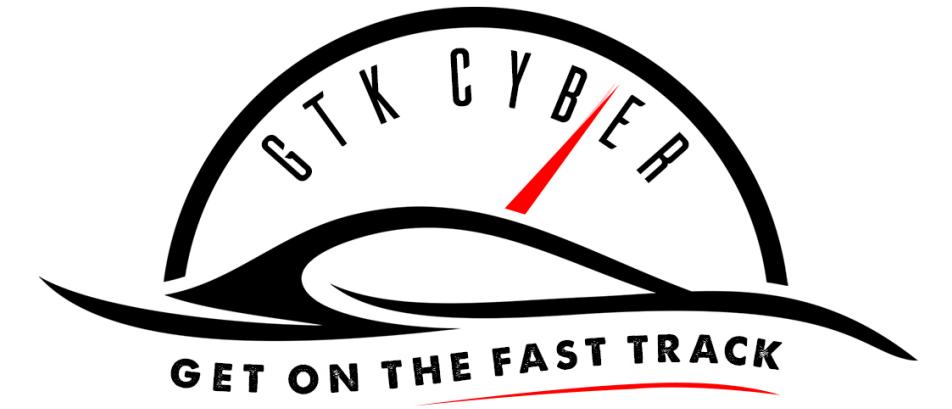




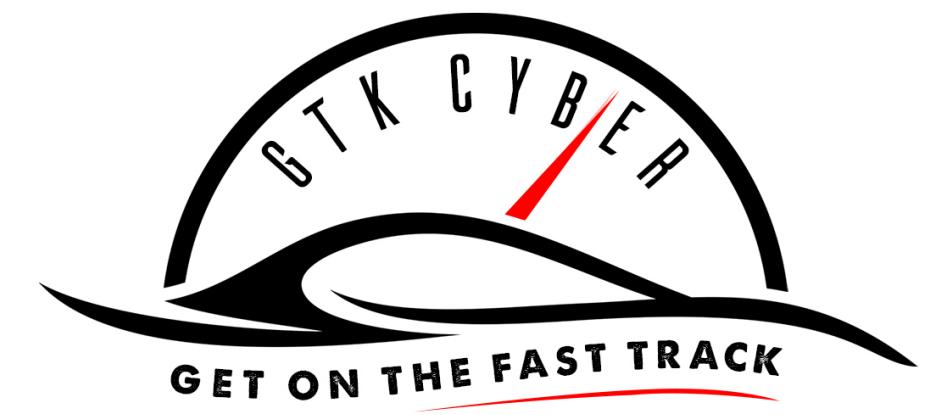


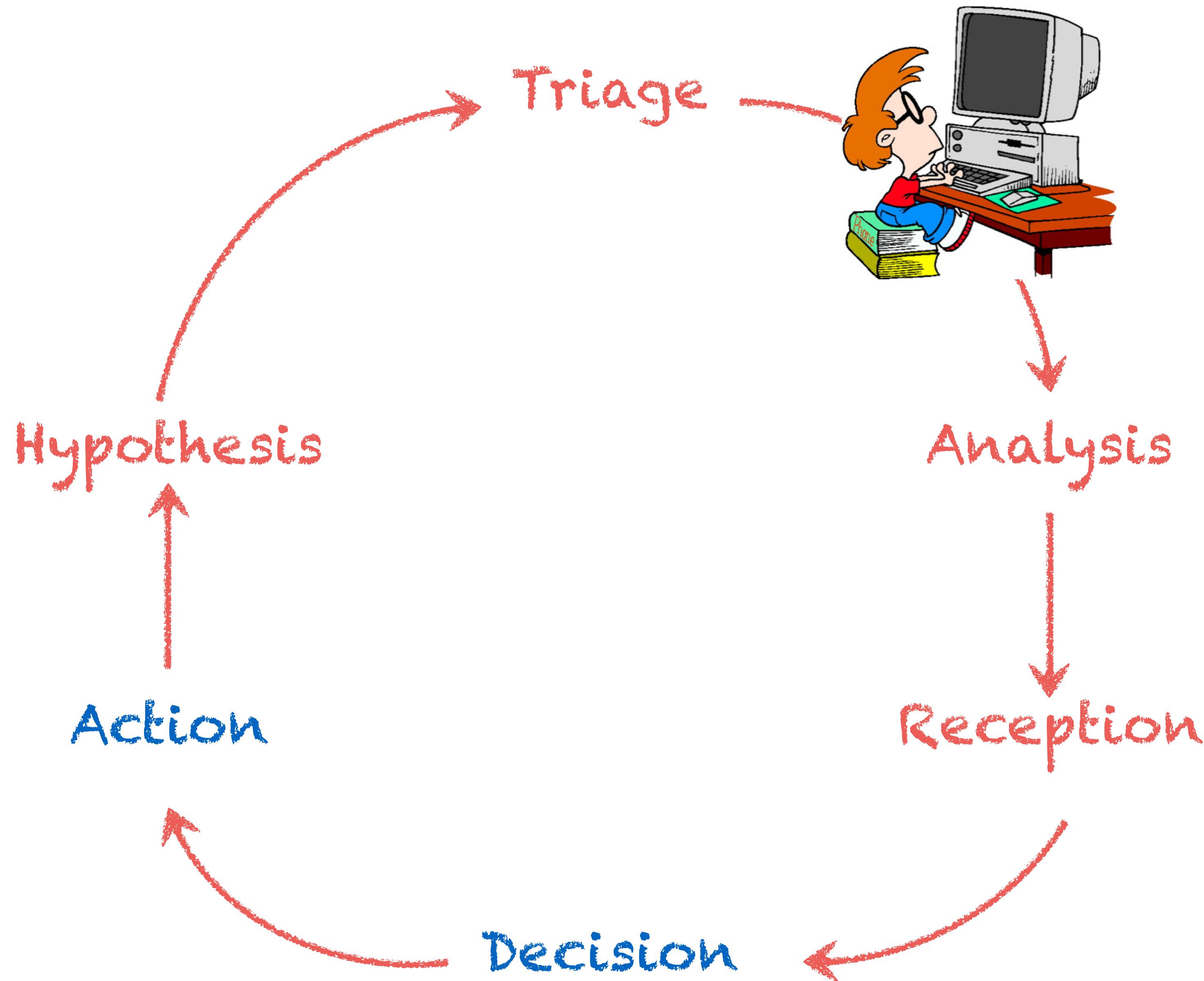
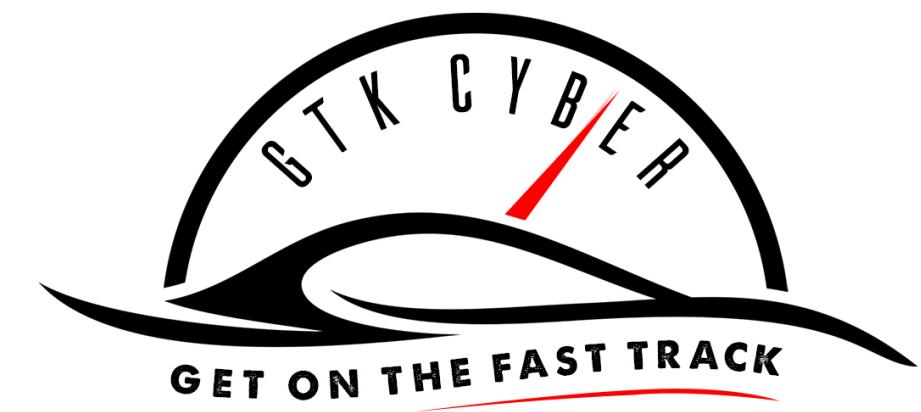
The Data Scientist Venn Diagram





Data Scientists spend
50-90% of their time being...







Thoughts for Data Science Success

Data is a Strategic Asset... not a cost



Align Projects to Corporate Strategy

Align Projects to Corporate Strategy



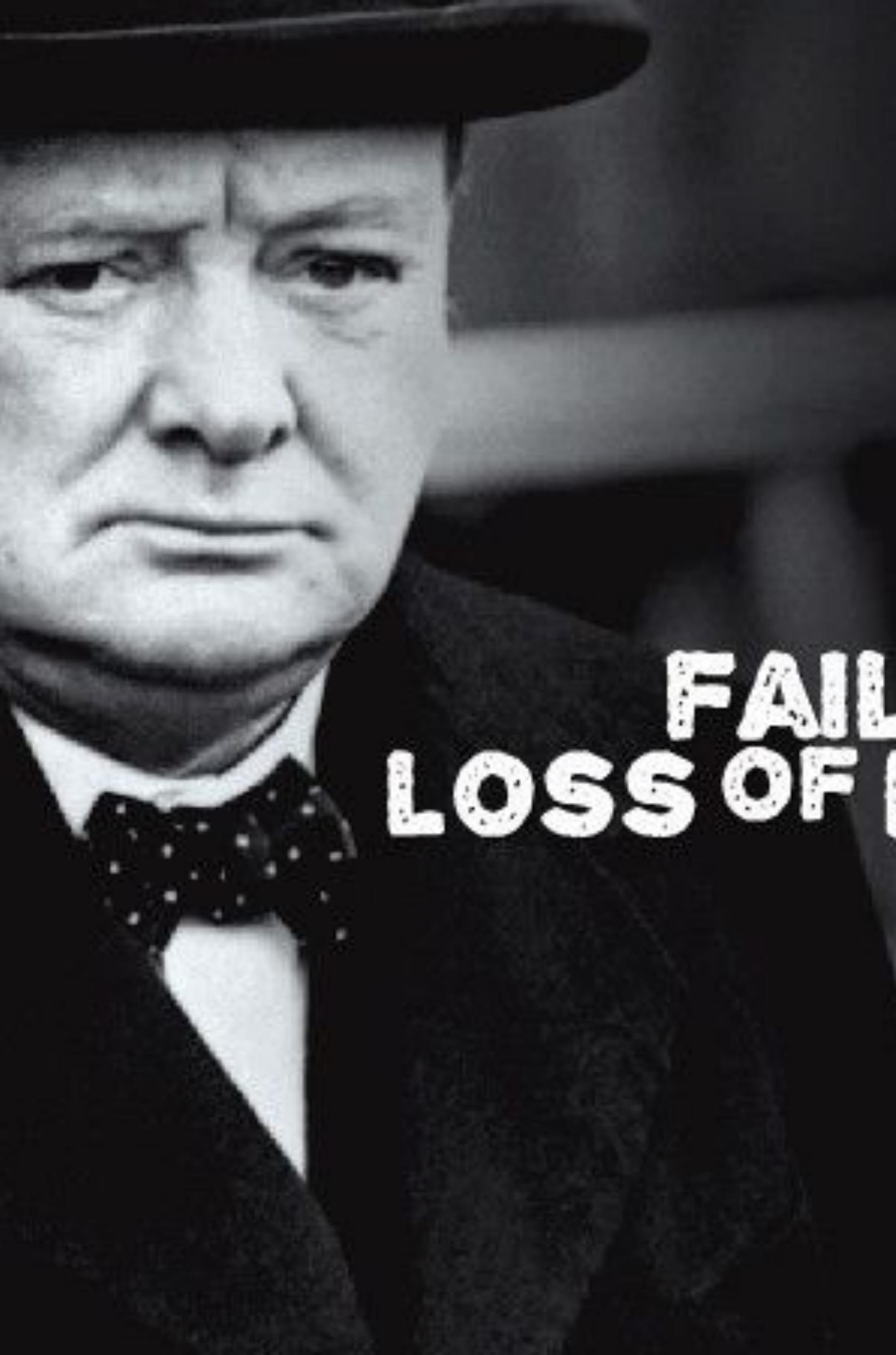
Your:
Time
Money
Job?

Build the right team for Data Initiatives



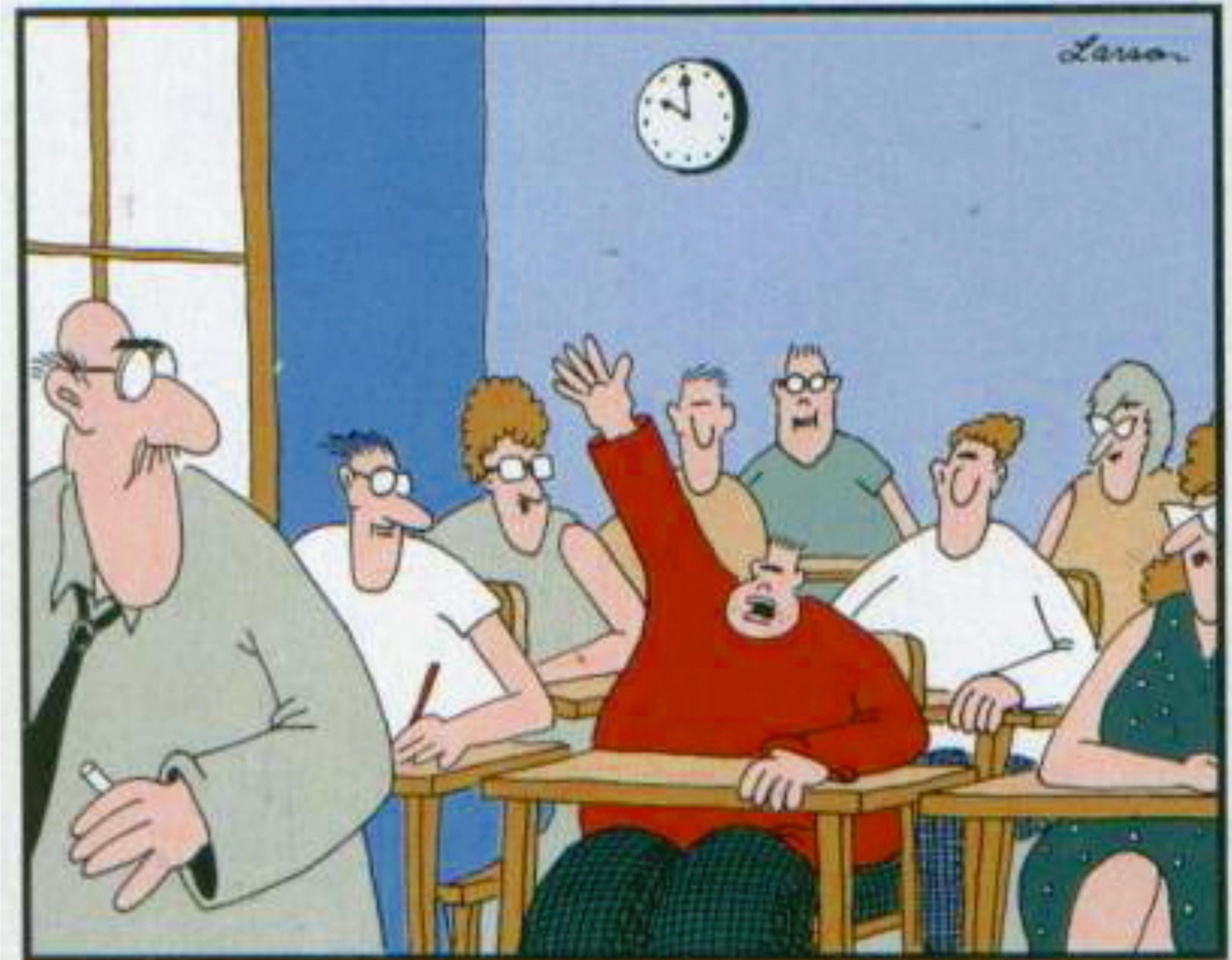
Prioritize building appropriate data platform



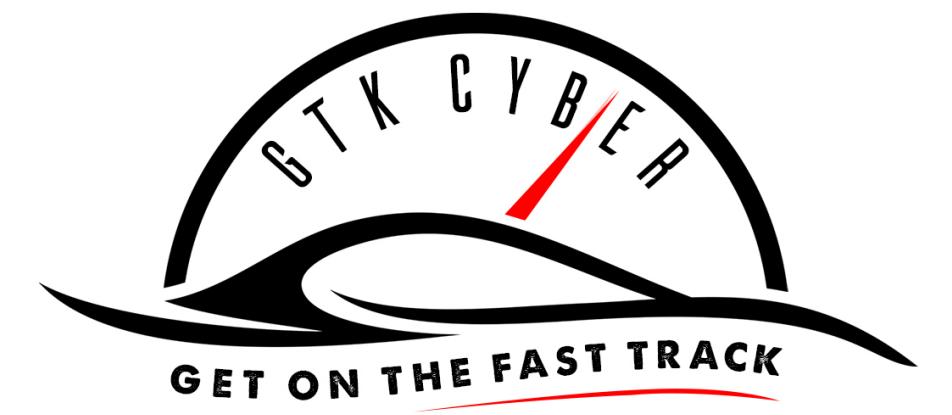


**"SUCCESS
CONSISTS OF
GOING FROM
FAILURE TO
FAILURE WITHOUT
LOSS OF ENTHUSIASM."**

Winston Churchill



**"Mr. Osborne, may I be excused?
My brain is full."**

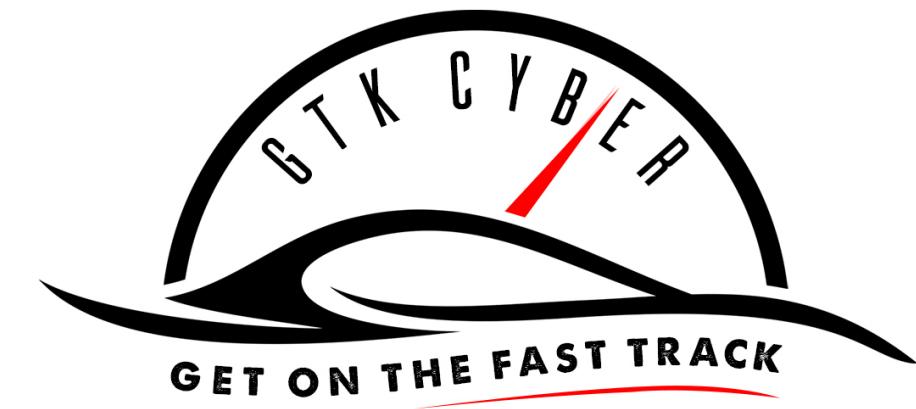


Building a Data Science Team



Building a Data Science Team

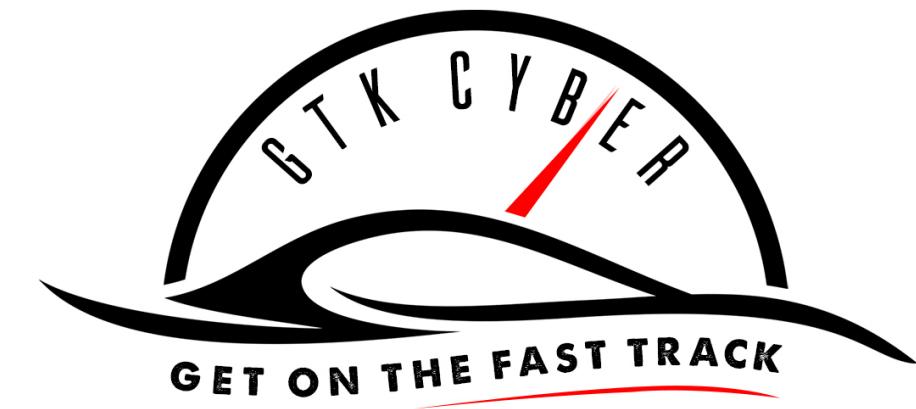
Programmer + Statistician + SME = DS Team?



Building a Data Science Team

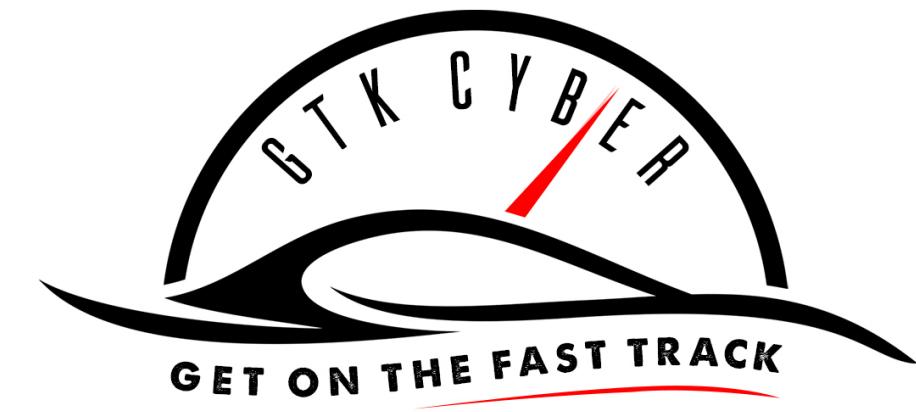
Programmer + Statistician + SME = DS Team?

Sort of...



Building a Data Science Team

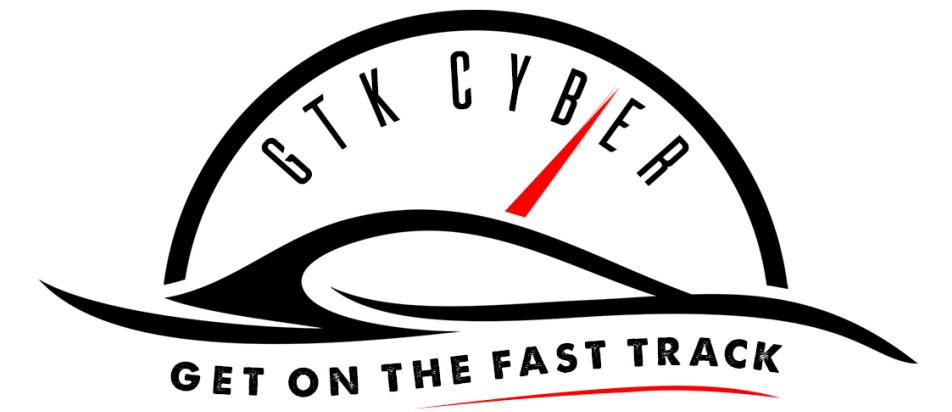
- Data Ambassador
- Data Scientist
- Data Storyteller
- Data Engineer

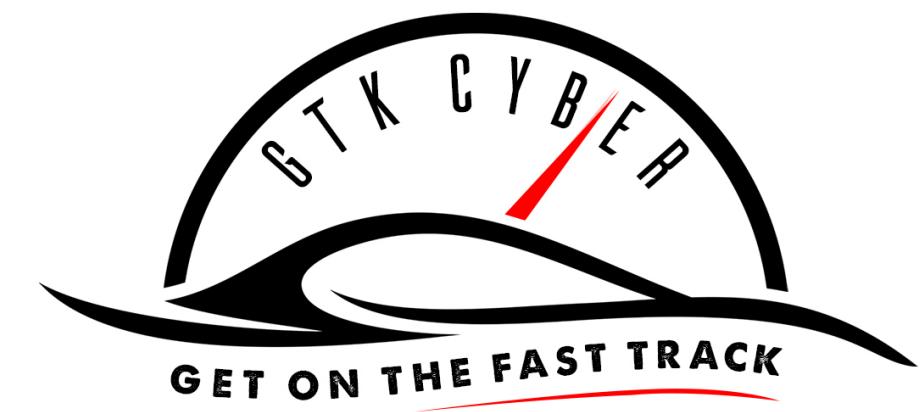


By the end of the class
you will be able to:

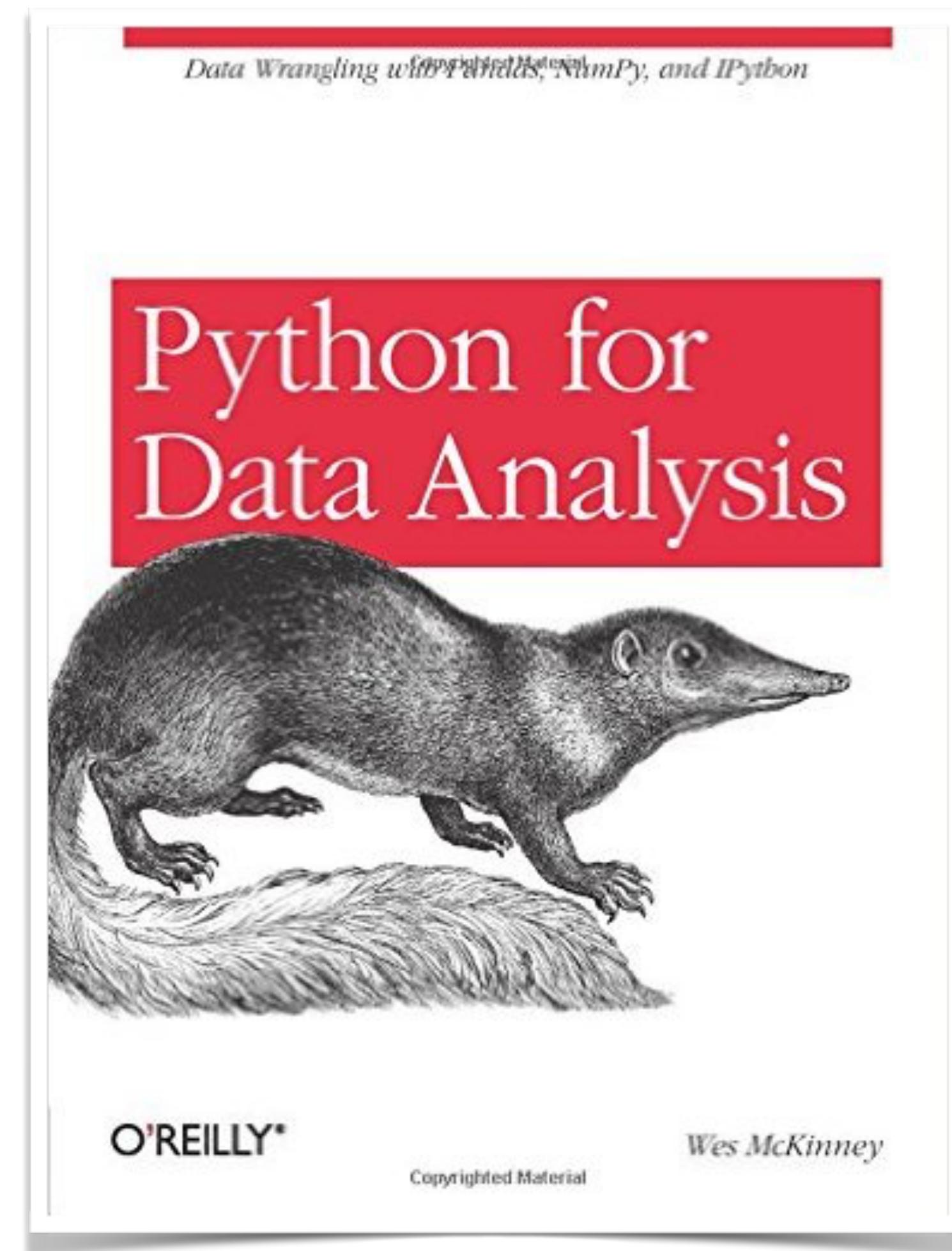
- Quickly and effectively prepare data for analysis
- Apply machine learning techniques to enhance security

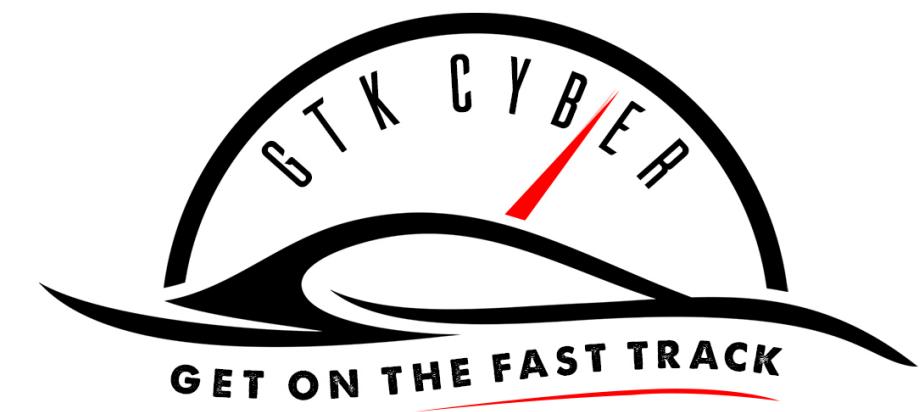




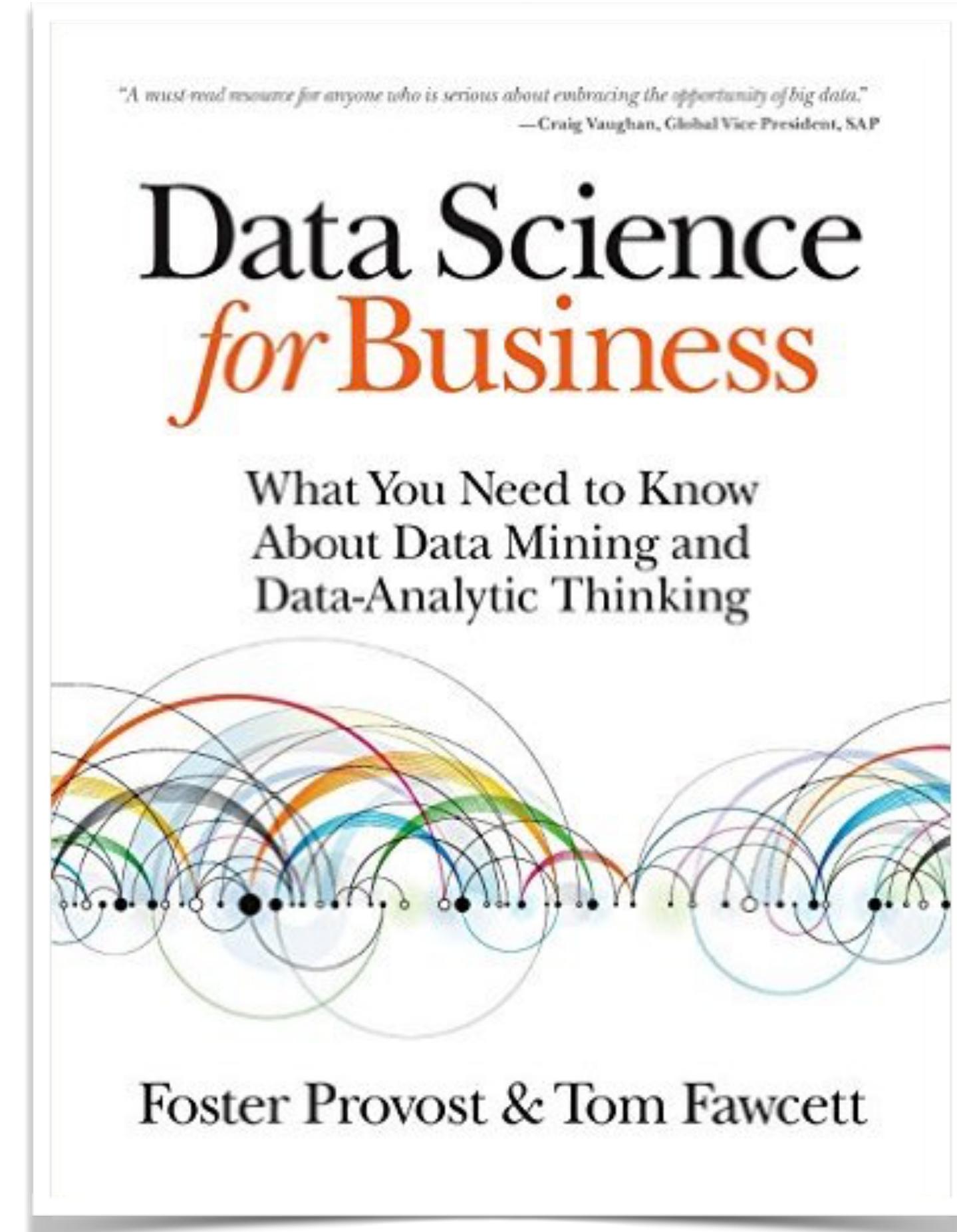


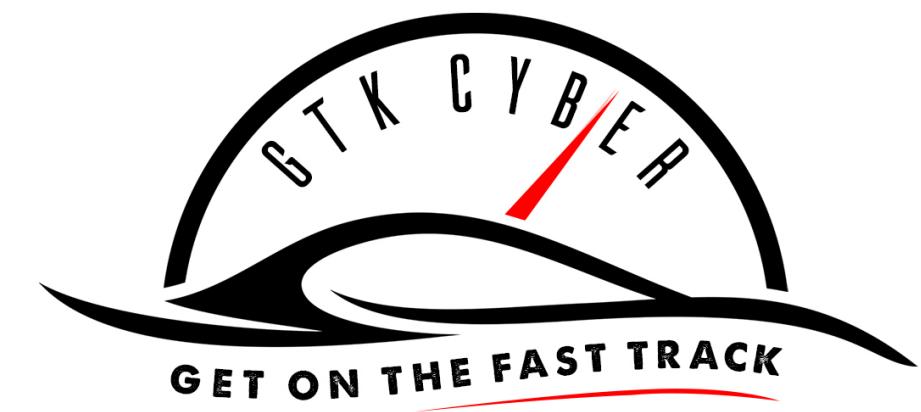
Recommended Reading



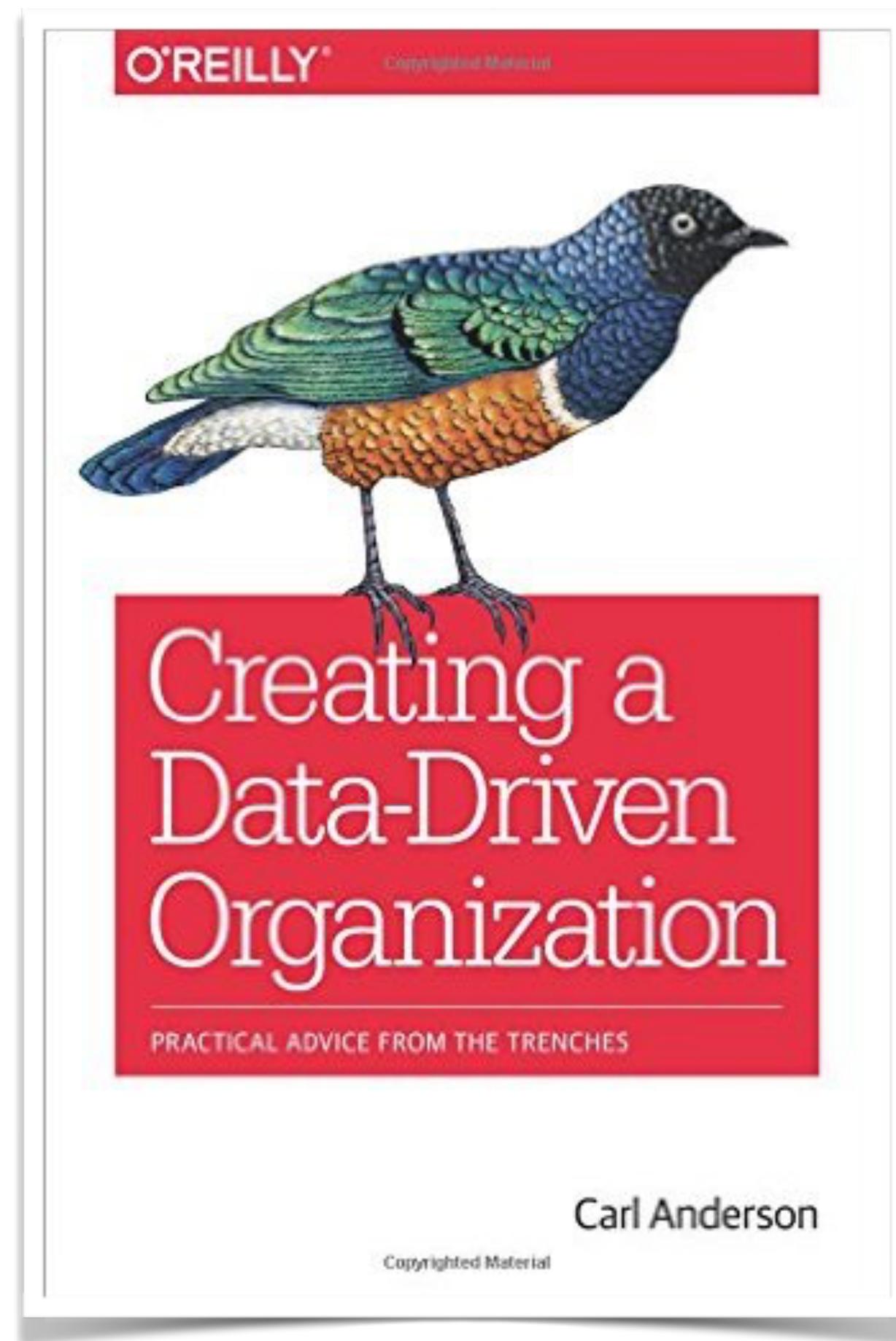


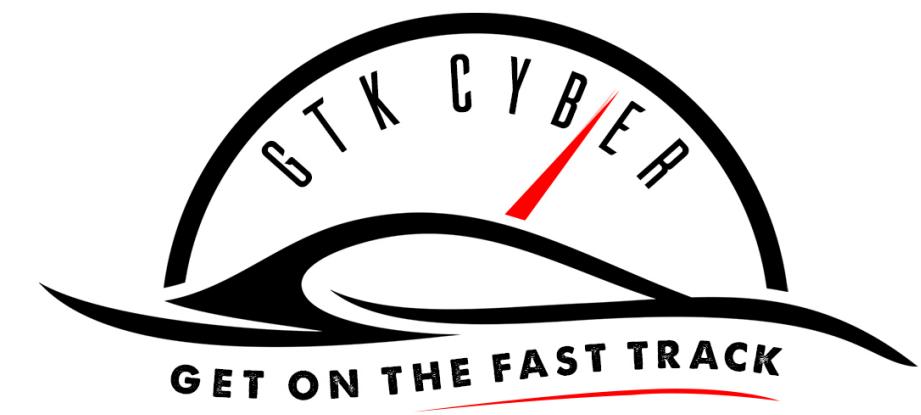
Recommended Reading



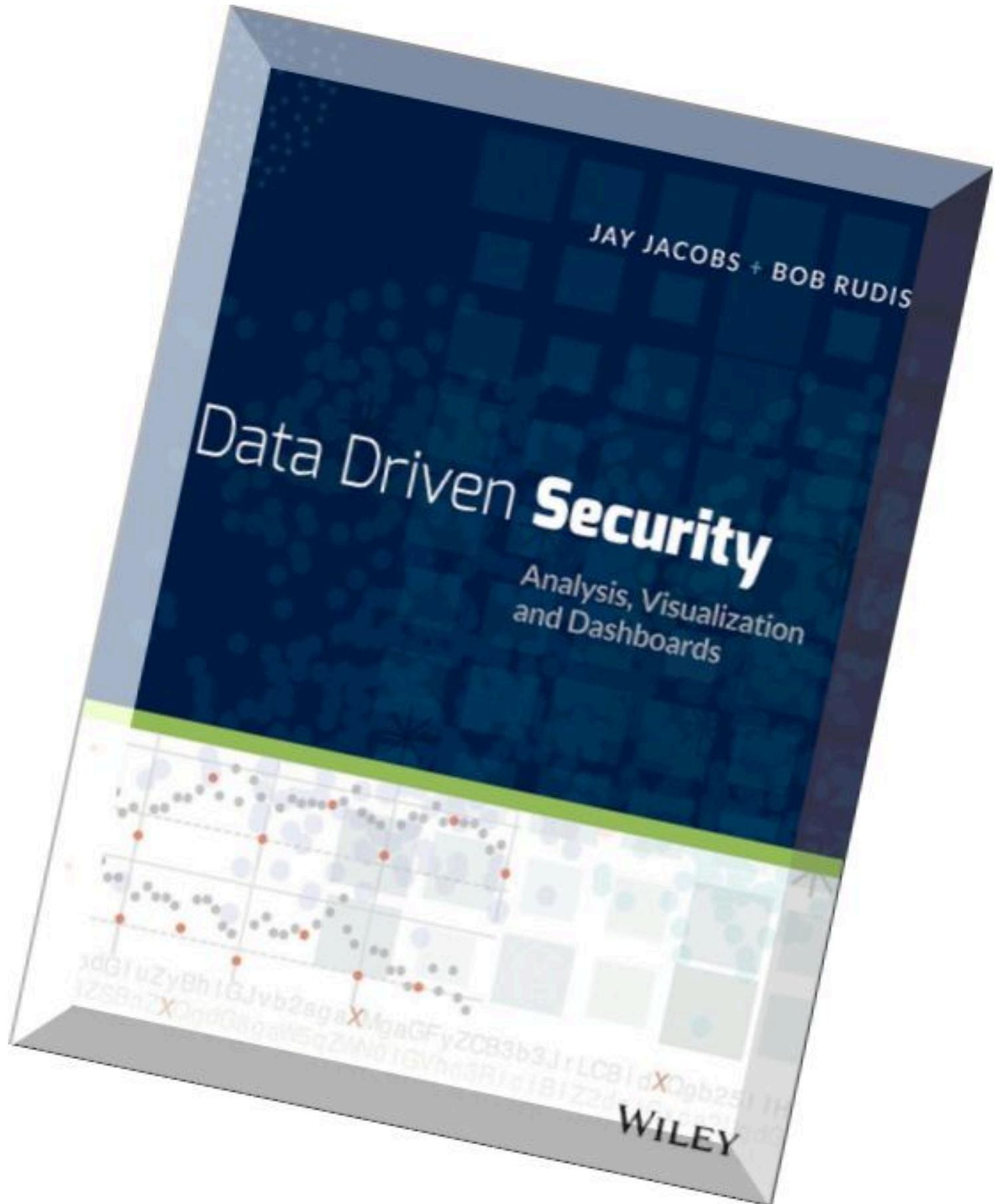


Recommended Reading



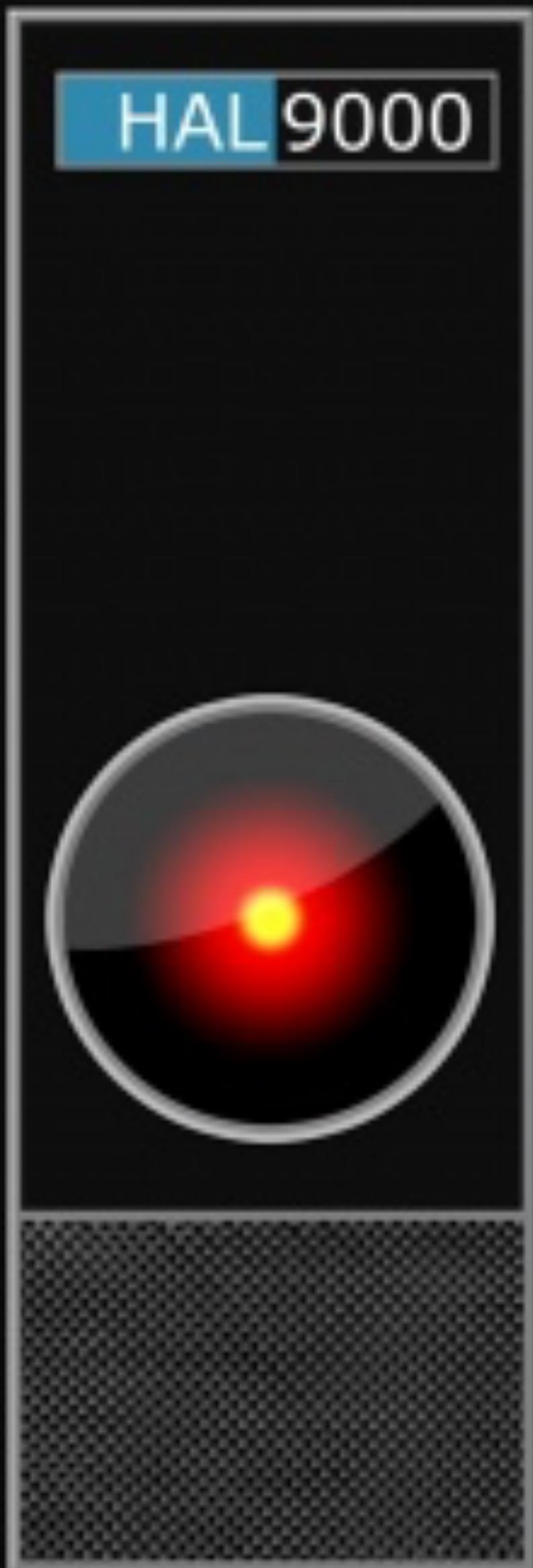


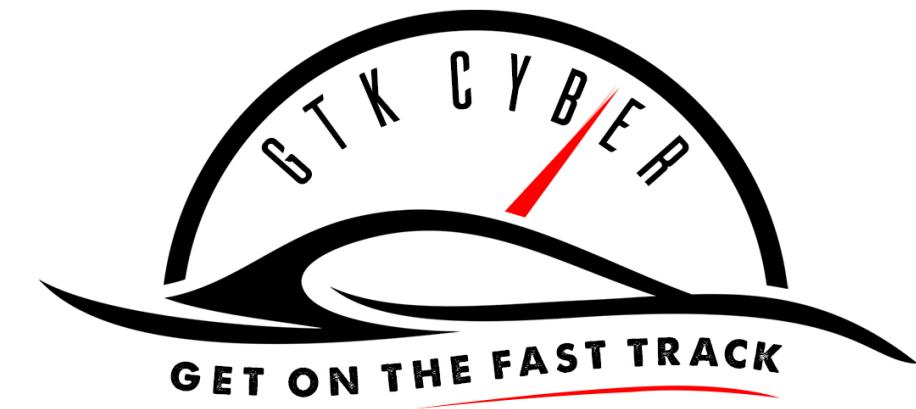
Recommended Reading



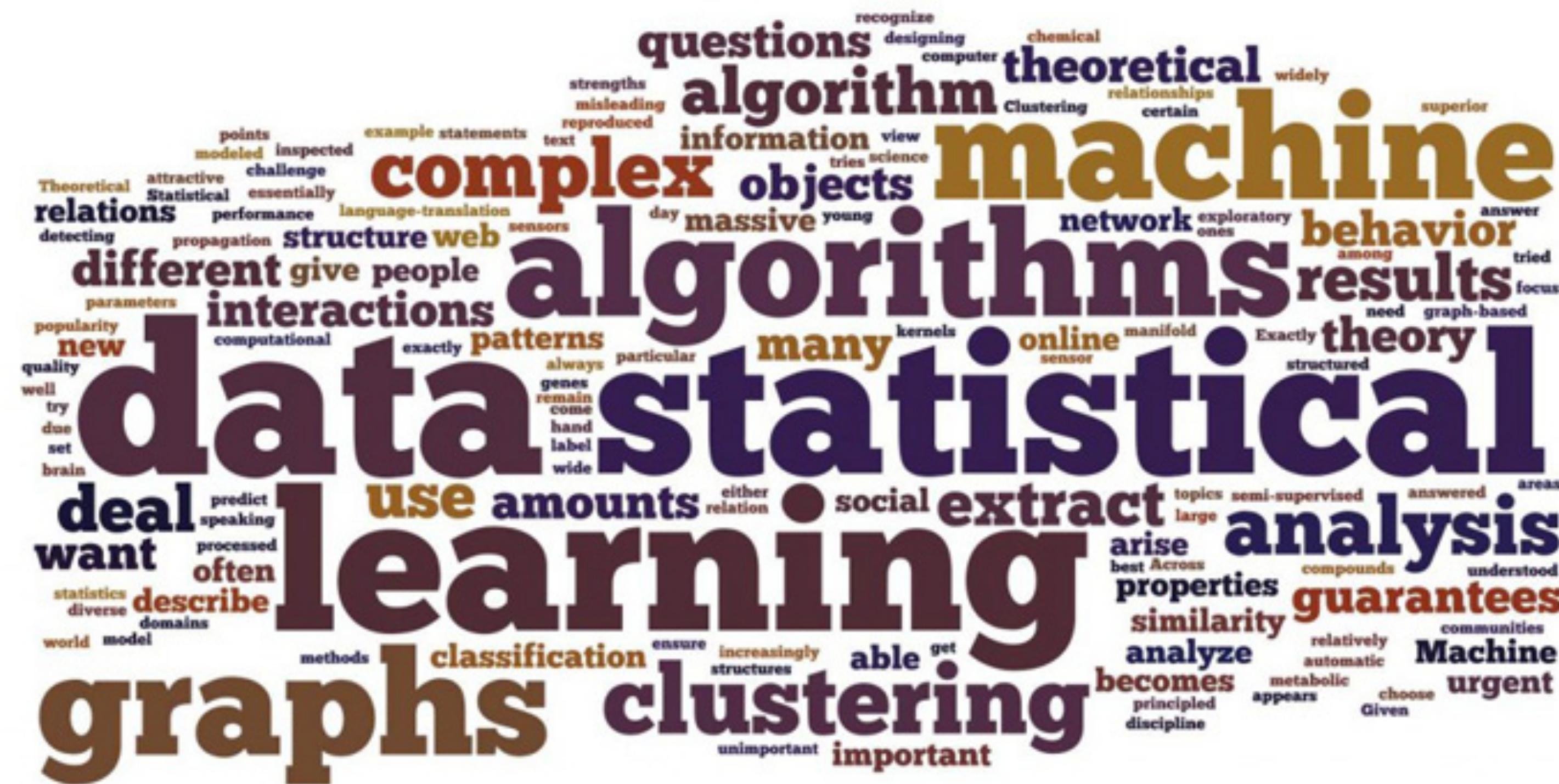
<http://datadrivensecurity.info>

Machine Learning



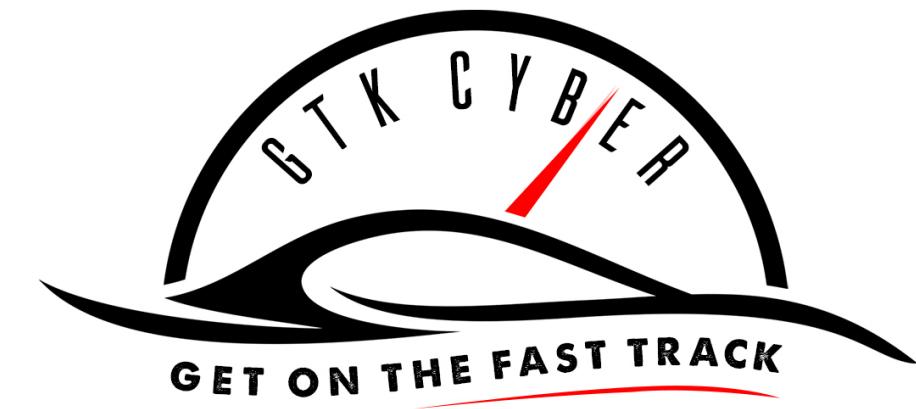


What is Machine Learning (ML) Artificial Intelligence (AI)



“I have no idea. The latest buzzword? ”

-The guy in the back of the class



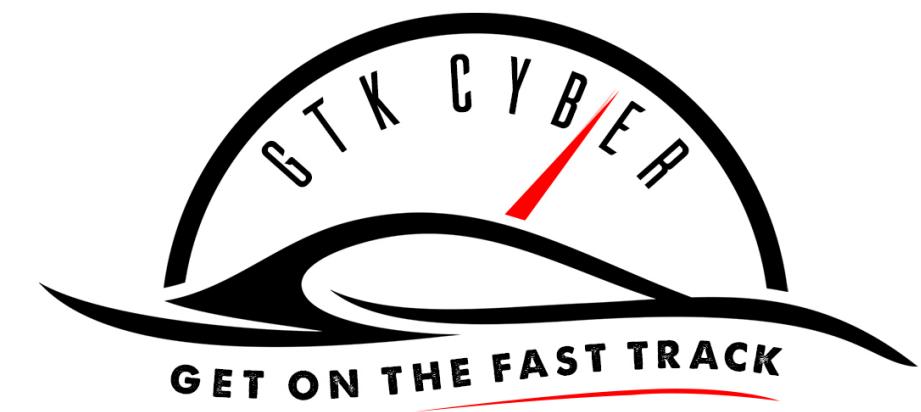
“Machine learning is the science of getting computers to **act without being explicitly programmed.** ”

– <https://www.coursera.org/course/ml>



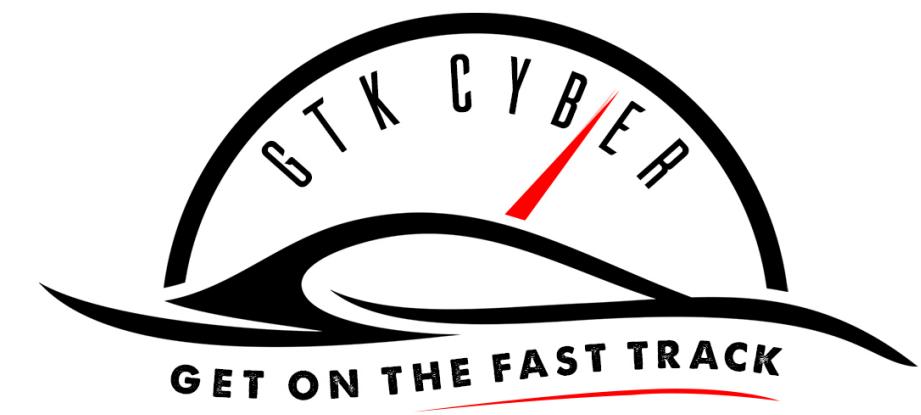
“A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”

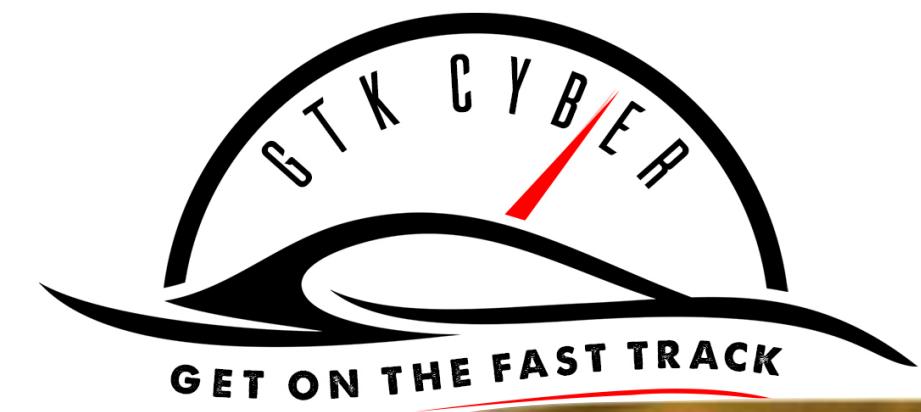
–Tom Mitchell, Carnegie Mellon University

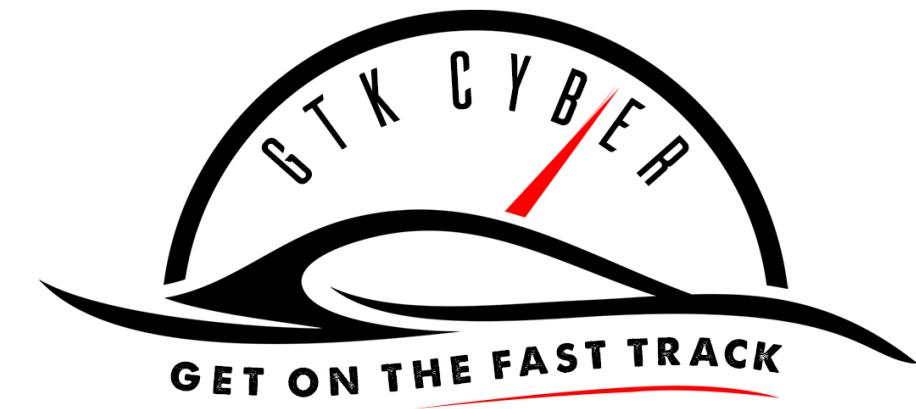


“Machine learning explores the construction and study of algorithms that can learn from and **make predictions on data**. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, **rather than following strictly static program instructions.**”

—https://en.wikipedia.org/wiki/Machine_learning







- Blacklists
- Simple keyword matching
- Naive Bayesian Classifiers
- Deep Learning



Artificial Intelligence

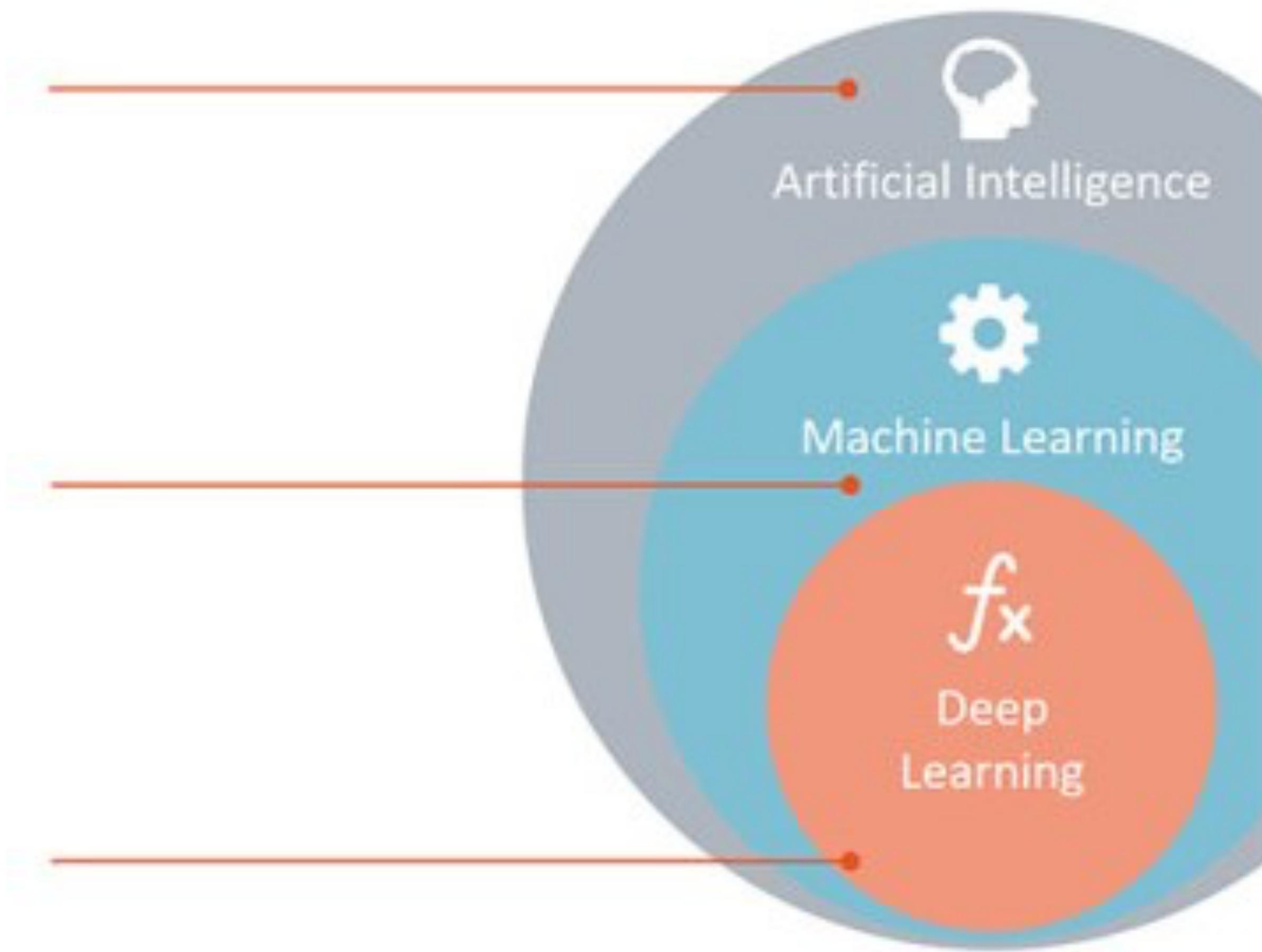
Any technique which enables computers to mimic human behavior.

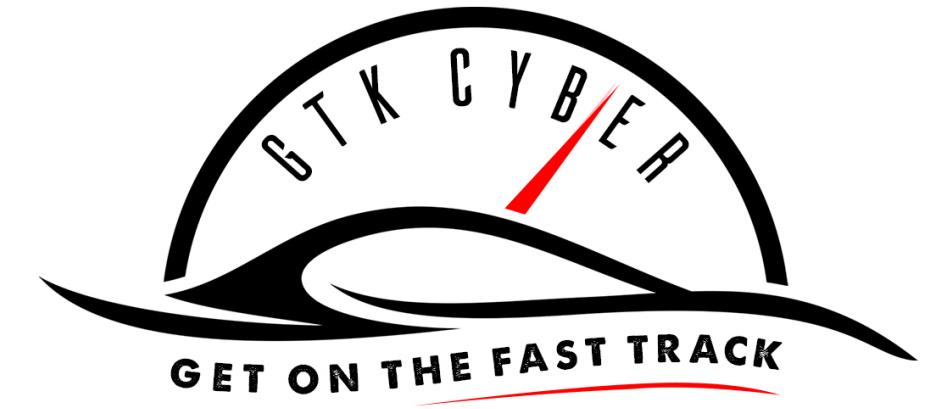
Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

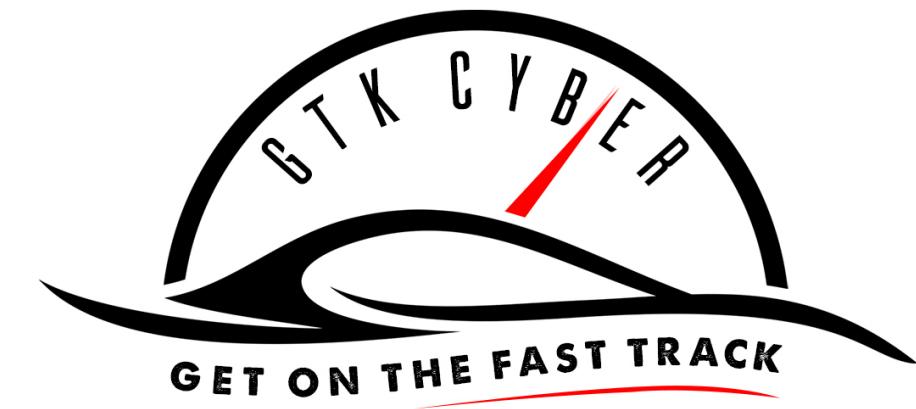
Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.



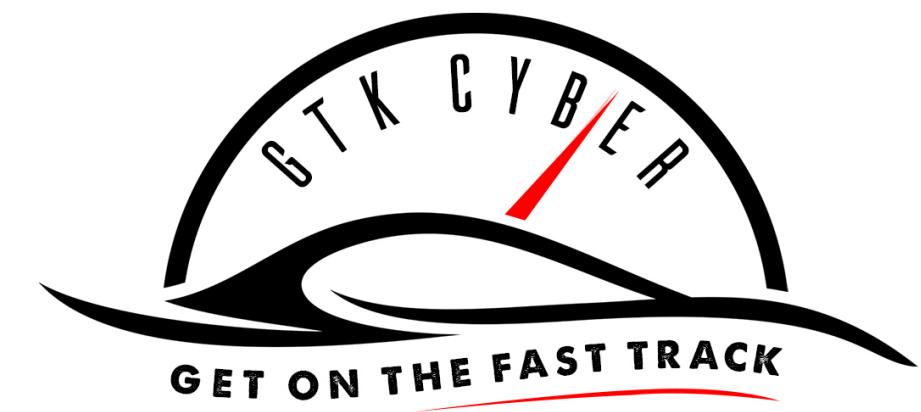


@katherinebailey Because marketing? Every time someone calls simple linear regression “AI” Gauss turns over in his grave.



Machine Learning Problems

- **Supervised Learning:** Supervised Learning is a class of Machine Learning in which a model is "trained" using a set of pre-existing labeled data.
- **Unsupervised Learning:** A class of Machine Learning algorithms in which a model is built without the use of labeled data.

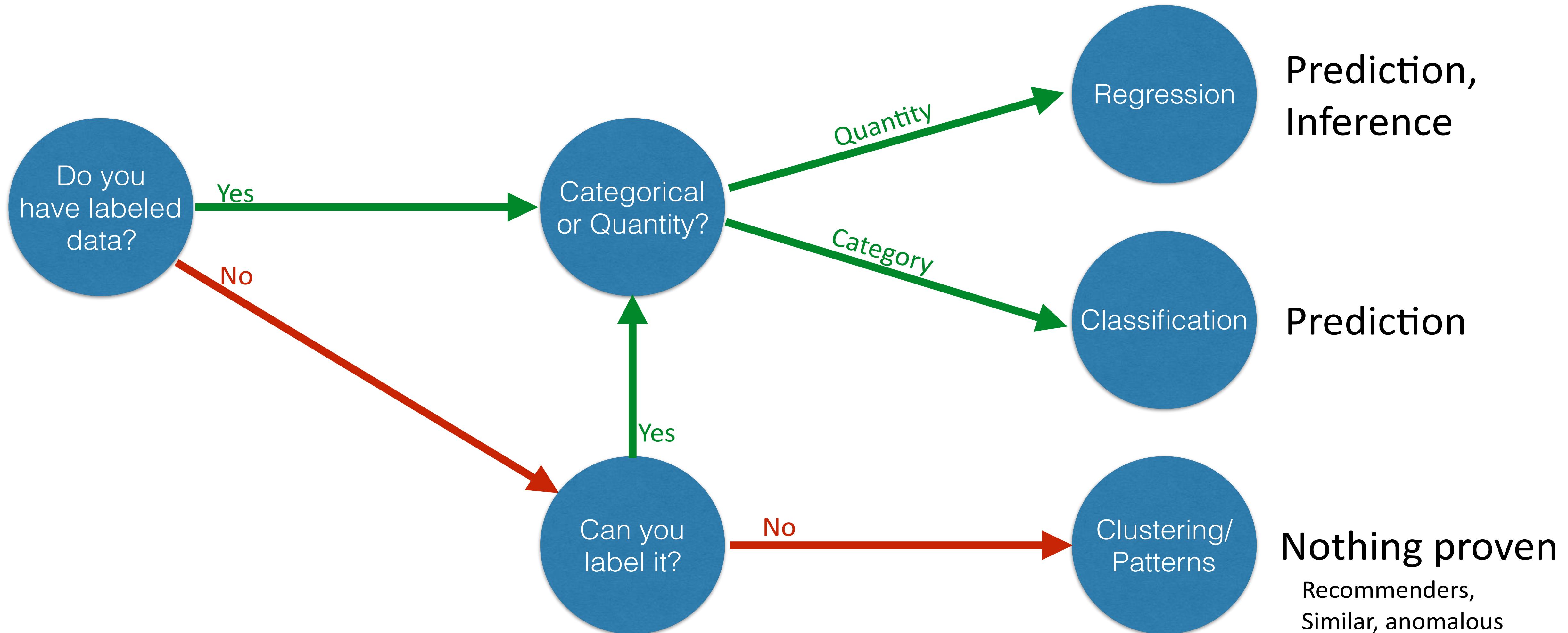


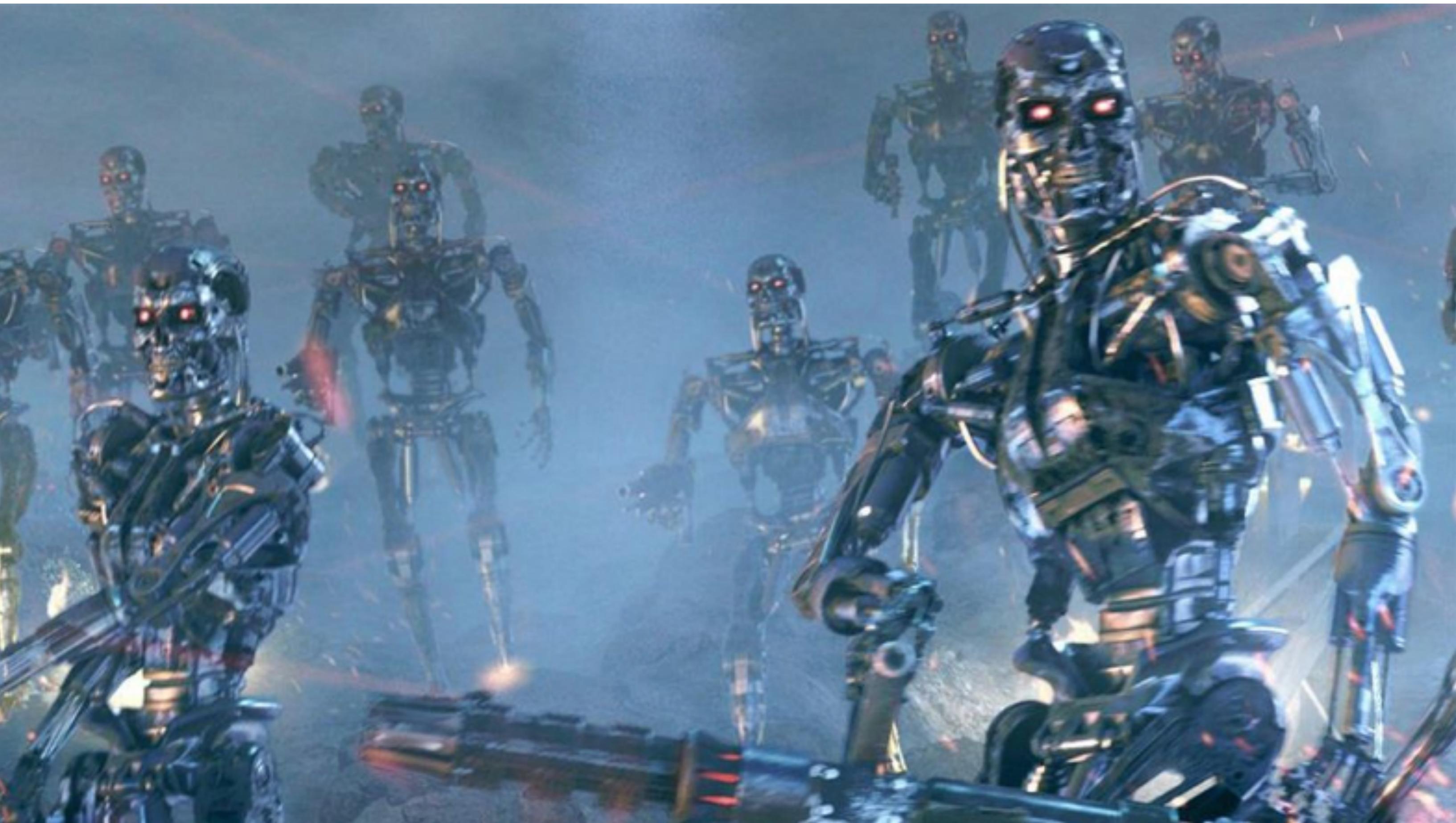
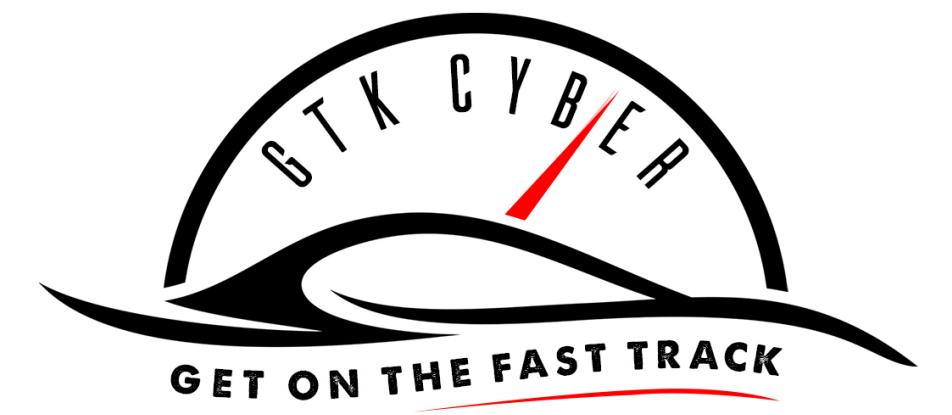
Machine Learning Problem Types

- **Classification:** Assigning or predicting a observation's membership in discrete class
- **Regression:** Predicting a continuous value based on the observations' features
- **Clustering:** Identifying groupings within a dataset
- **Dimensionality Reduction:** Reducing the number of variables in a feature set

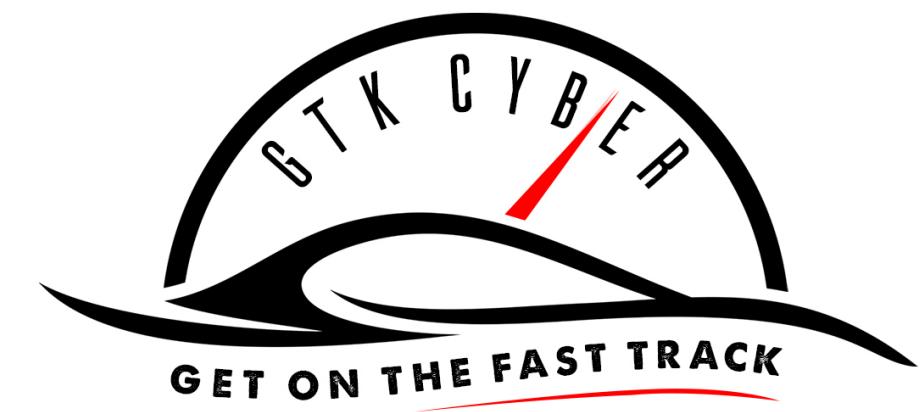


What Problem am I solving?

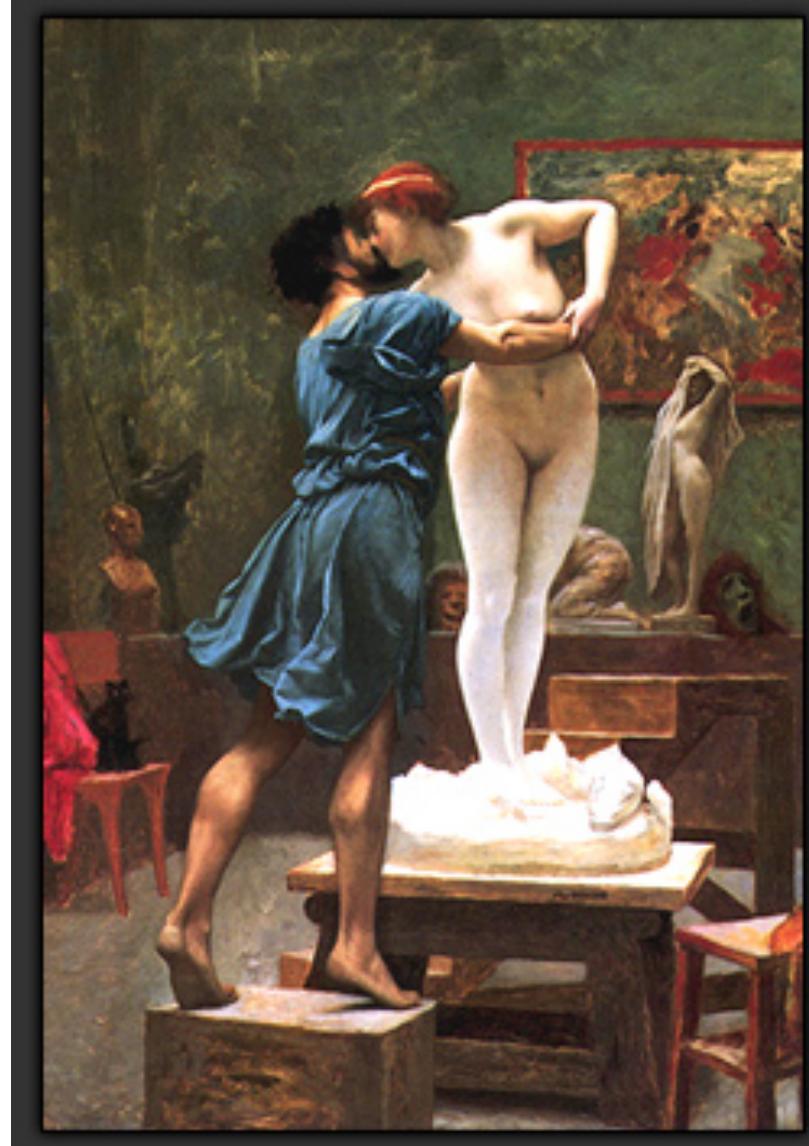




What it is not

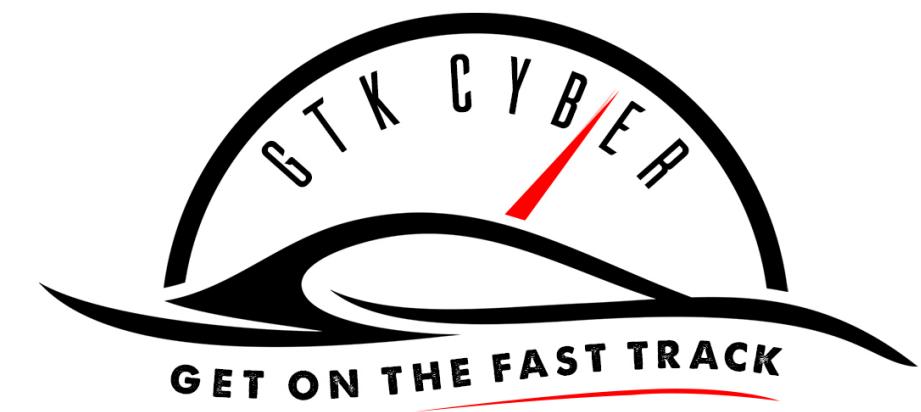


History of AI and ML



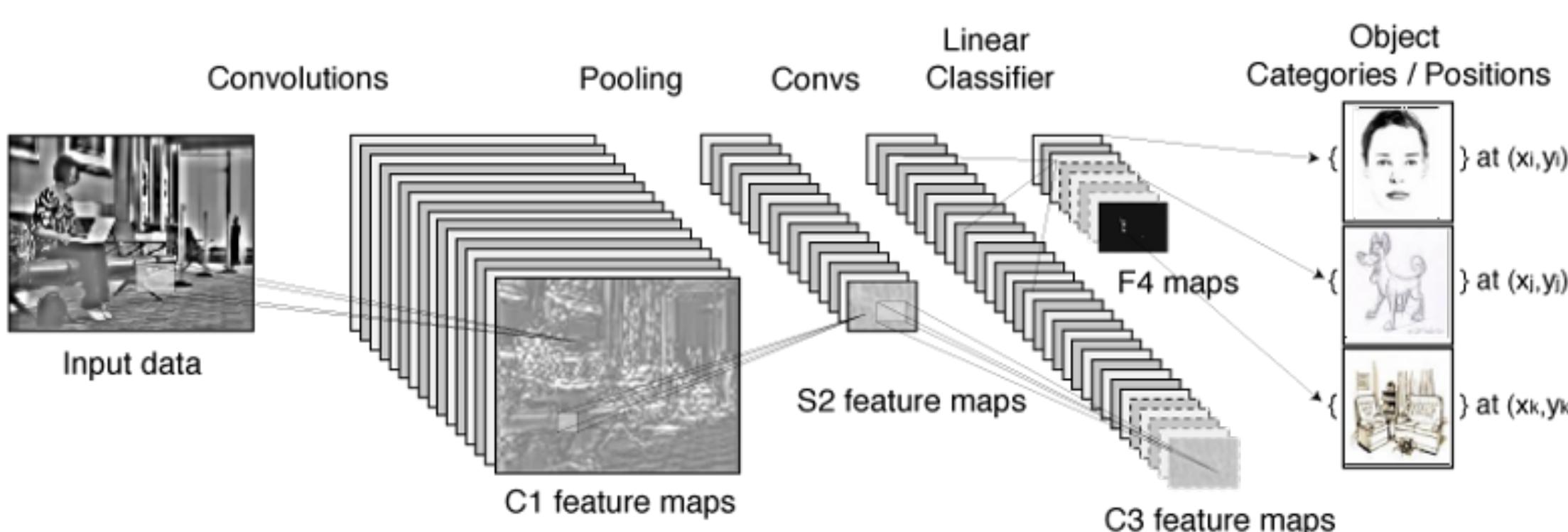
- Mythical figure Pygmalion: Breath of life in a statue (ancient Greece)
- Turing Test 1950s by Alan Turing: Test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.
- Early AI: solve problems that are difficult for humans, but straightforward for computers (e.g. formal mathematical rules) IBM Deep Blue chess-playing system defeated world champion Garry Kasparov 1997.
- Real Challenge: Intuitive, easy problems for humans. They are hard to describe formally (e.g recognizing faces, words).

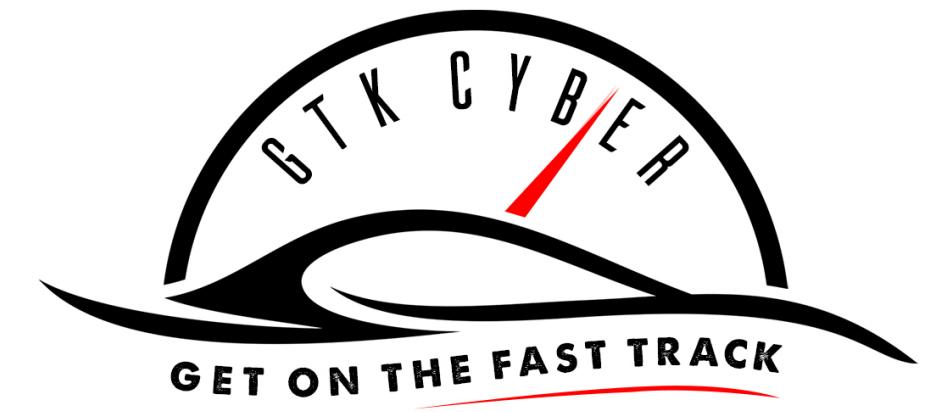




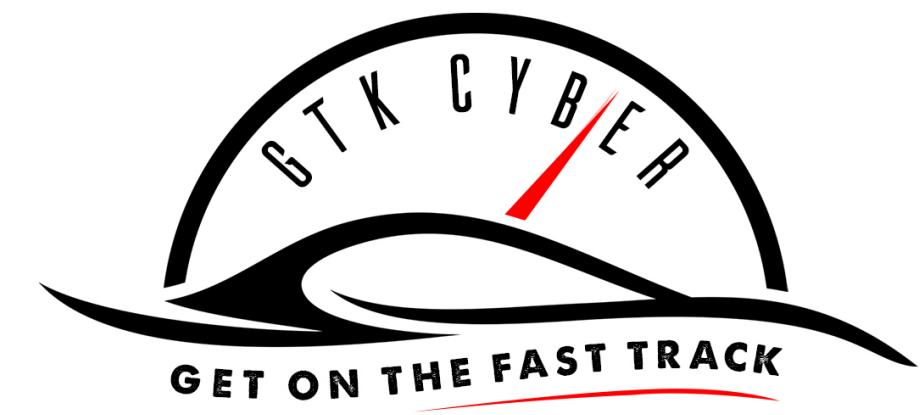
Path to Artificial Intelligence?

- Hard-coding knowledge failed (e.g. Cyc 1989).
- **Machine Learning: Gathering knowledge from experience** eliminates the need to formally specify all knowledge the computer needs.
- **AI** needs to acquire own knowledge by **extracting pattern from raw data.**
- Success depends heavily on **representation of data, aka features.**





Applications to Security

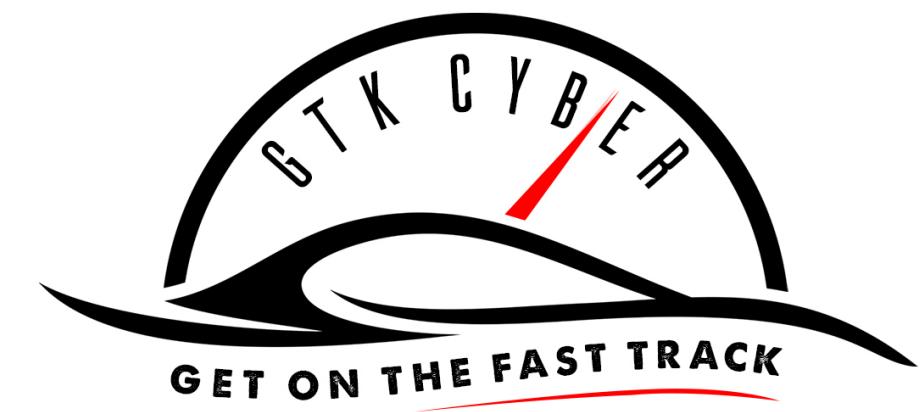


Regression Example

Server Capacity Prediction: Regression analysis can be used to predict a server's capacity (or CPU usage) based on the server's historical performance.

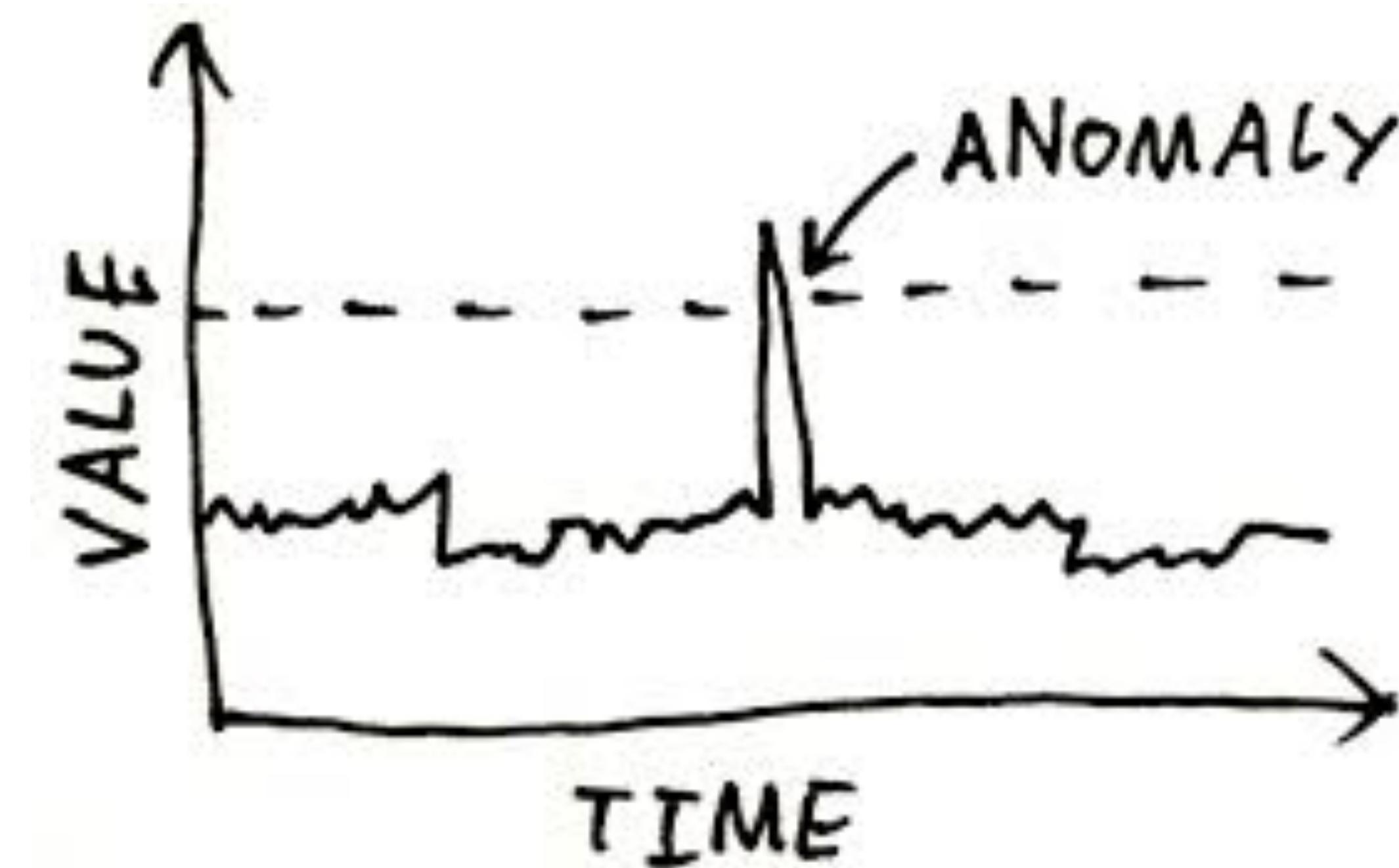


https://www.researchgate.net/publication/256645877_LiRCUP_Linear_Regression_based_CPU_Usage_Prediction_Algorithm_for_Live_Migration_of_Virtual_Machines_in_Data_Centers



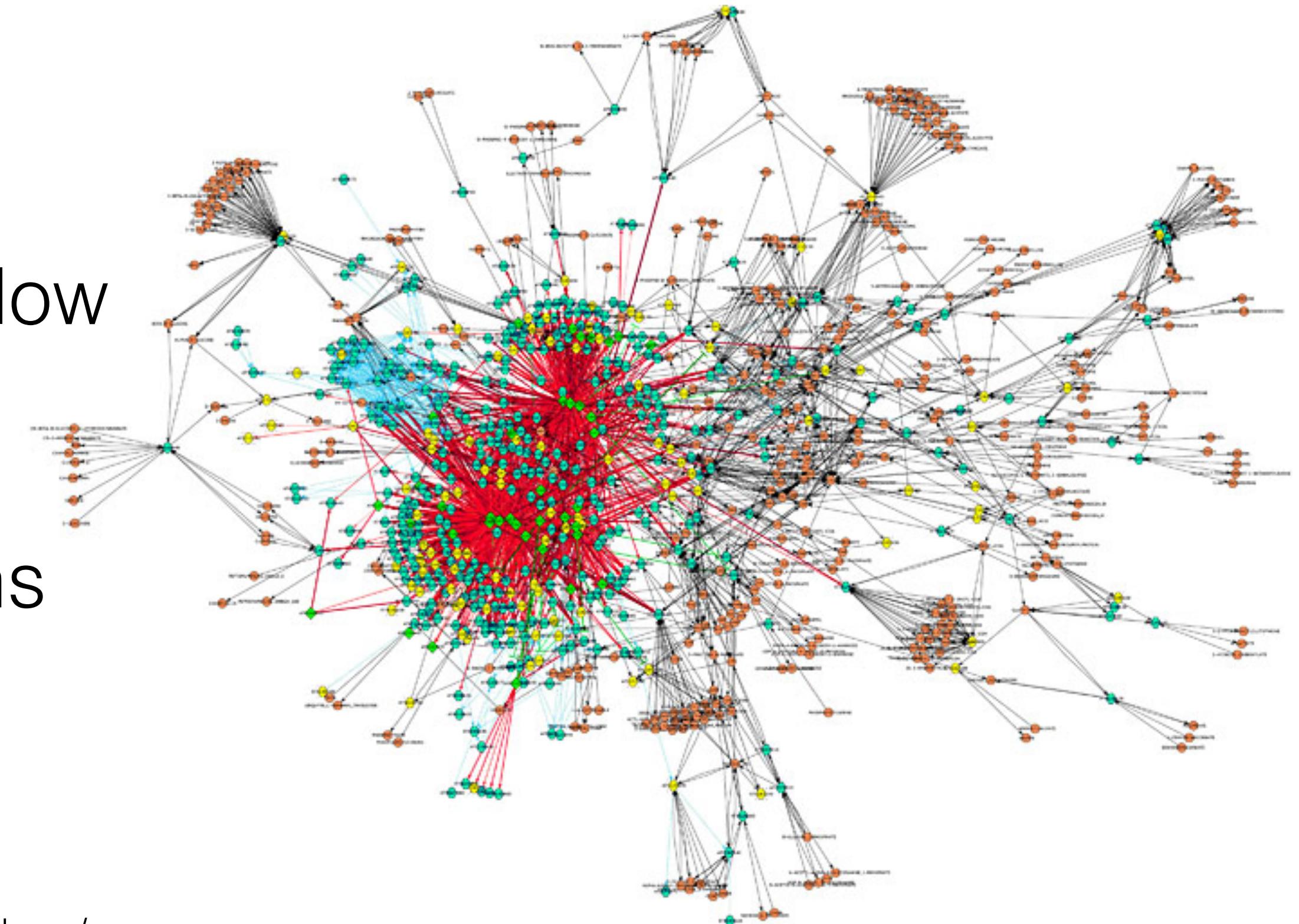
Clustering Example

Anomaly Detection: Clustering techniques can be used to detect anomalous traffic or loads or anything really.



Network-Based Intrusion Detection

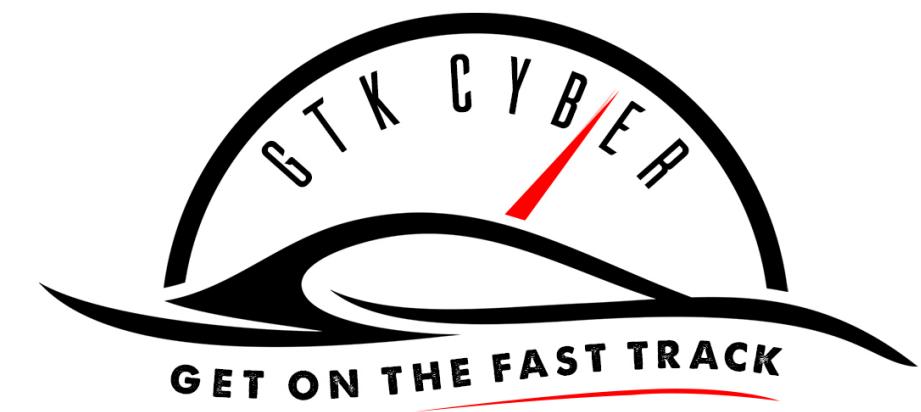
- Derive Features from Network Traffic
Captures “pcap” at packet level or NetFlow level (tools: tshark, tcpdump, bro...)
- Example Features based on header information: 2s-windowing of connections
> duration, protocol, src and dat bytes, service.
- Get data sets: <http://www.netresec.com/?page=PcapFiles>, <https://maccdc.org/>, <http://www.westpoint.edu/crc/SitePages/DataSets.aspx> <http://www.unb.ca/cic/research/datasets/>,



Malware Detection/Classification

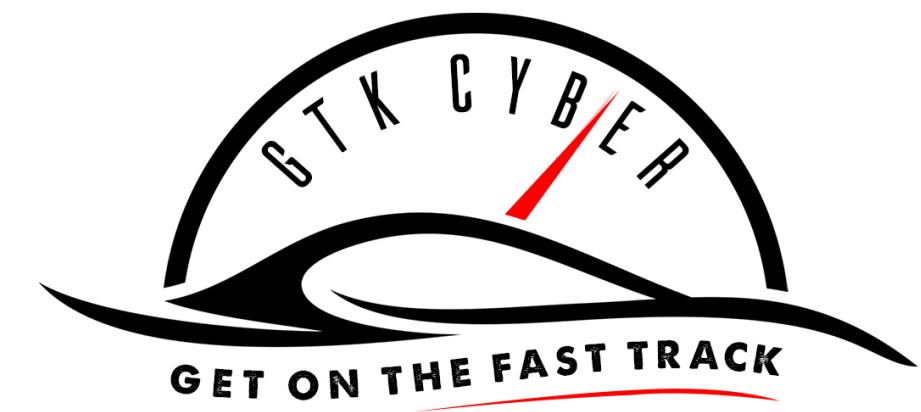
- Derive Features from Binary Content and metadata manifest (function calls, string obtained from IDA Disassembler)
- Example Features: opcode count (n-grams), segment count, asm pixel intensity, n-gramming of bytes, function name.
- Featureless Deep Learning with word2vec embedding
- Get open source malware samples: Vx Heaven, Virus Share, Maltrieve, Open Malware



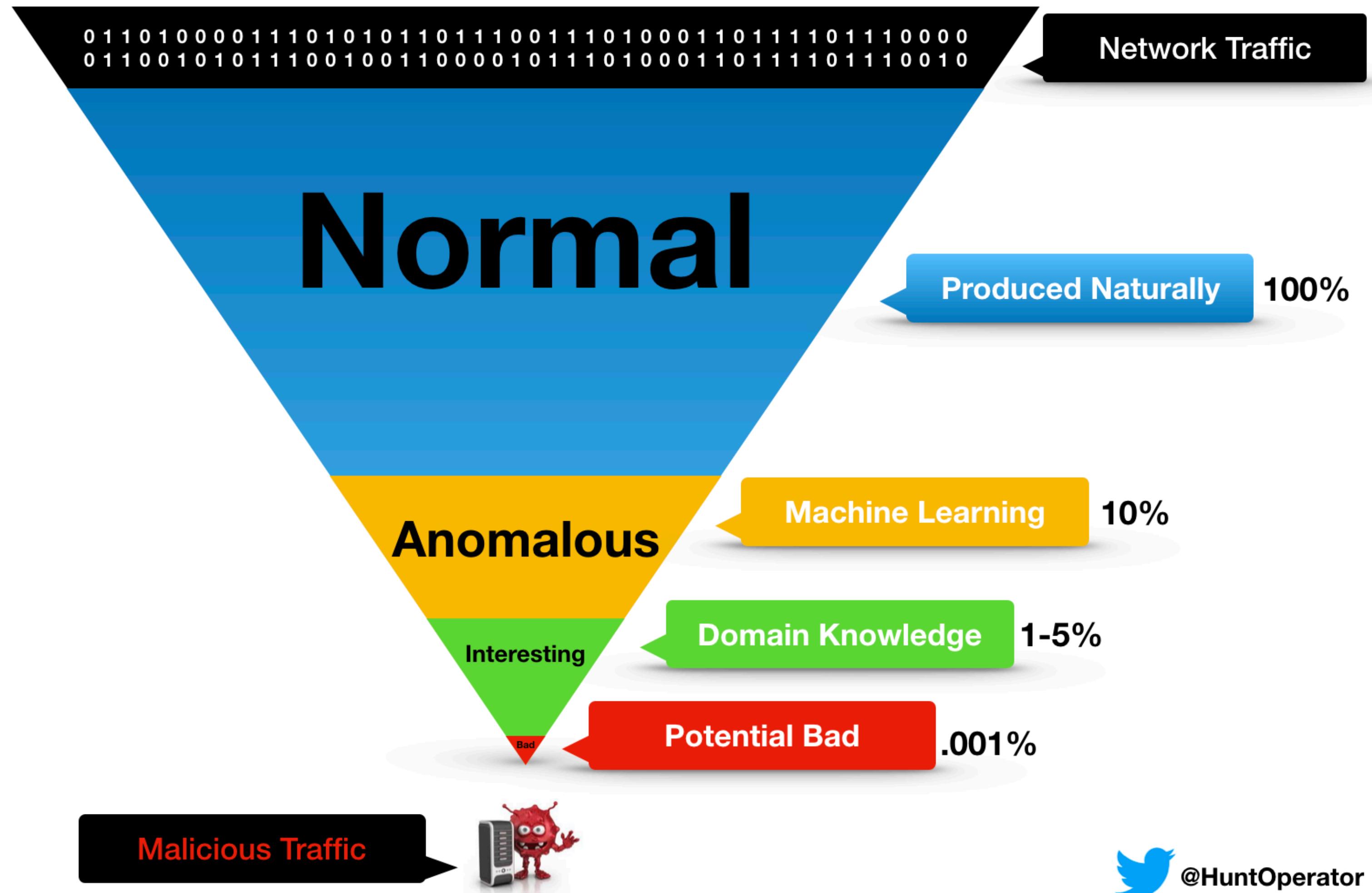


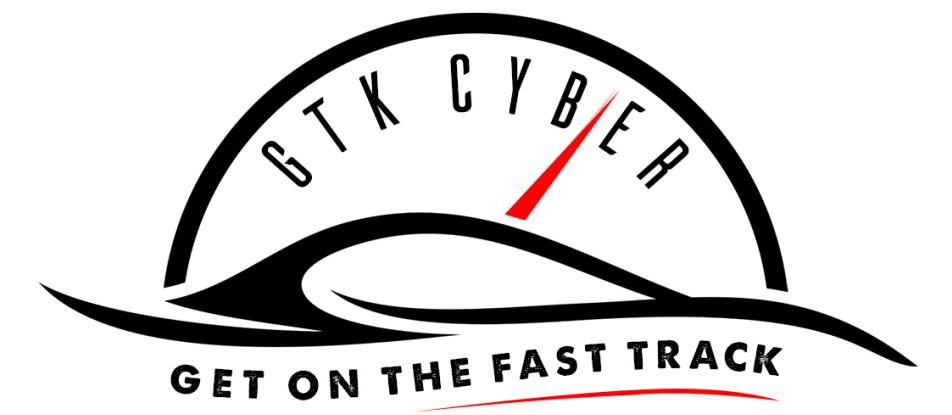
Security Applications of Machine Learning

- Domain Generation Algorithm (DGA) Detection (Classification)
- Malicious URL Detection (Classification)
- Network Traffic: Beaconing Detection (Classification/Clustering)
- Detection of new classes of malware (Classification/Clustering)
- General Network Traffic Anomaly Detection (Classification/Clustering)
- Log Analysis - Anomaly Detection (Classification/Clustering)
- Phishing Detection (Classification)
- Identifying SQL Injection (Classification)
- Identifying XSS cross-site scripting (Classification)
- DOS/DDOS Detection (Classification)
- Authentication (Classification)



Data Science Hunting Funnel

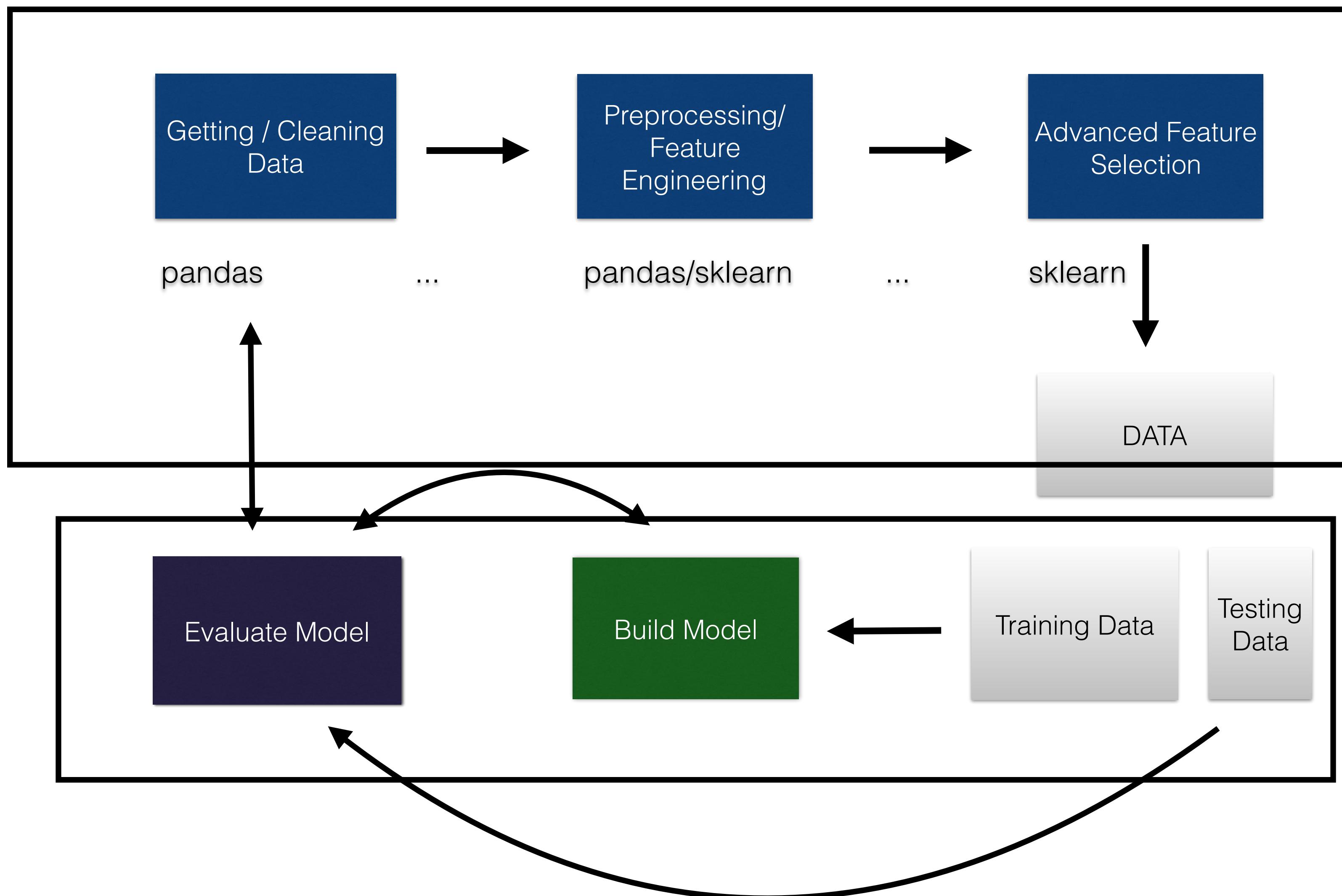


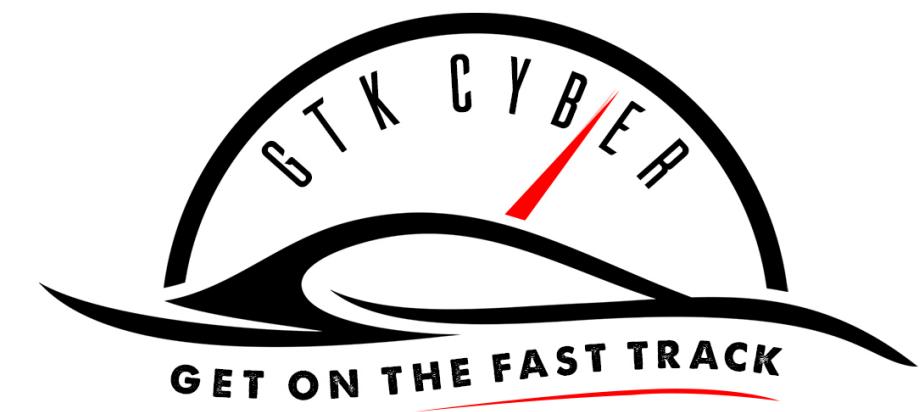


The Machine Learning Process

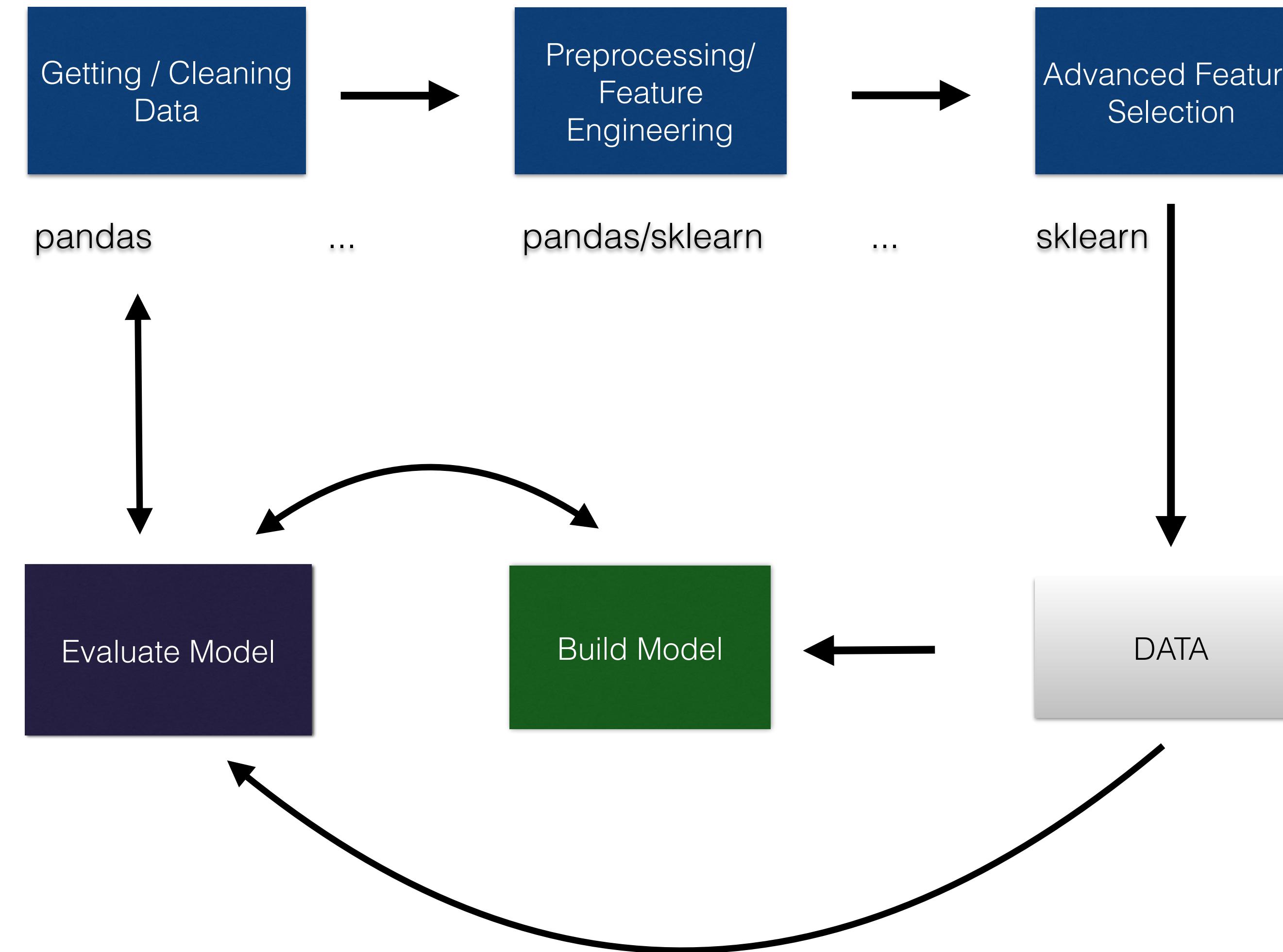


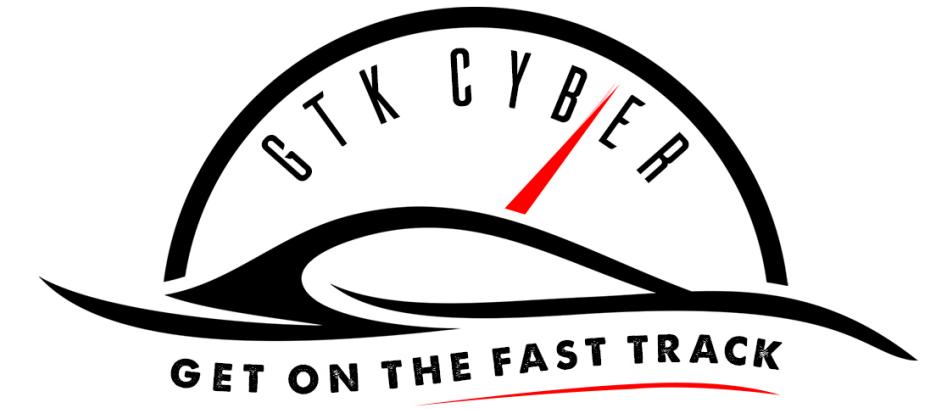
Supervised Machine Learning Process



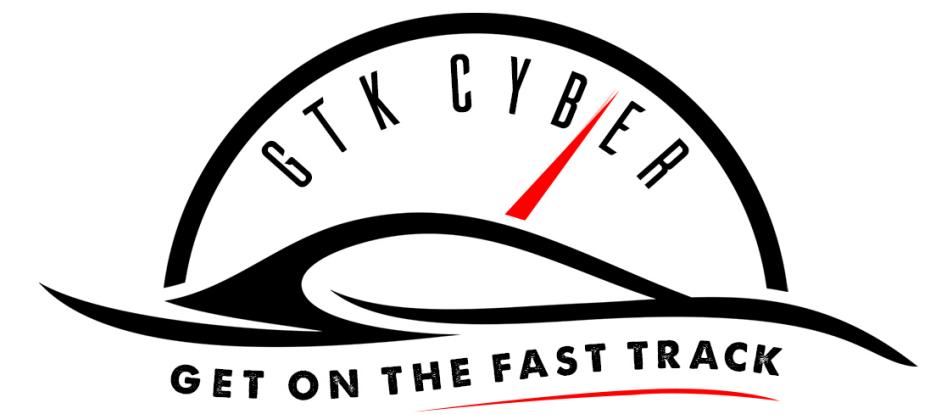


Unsupervised Machine Learning Process

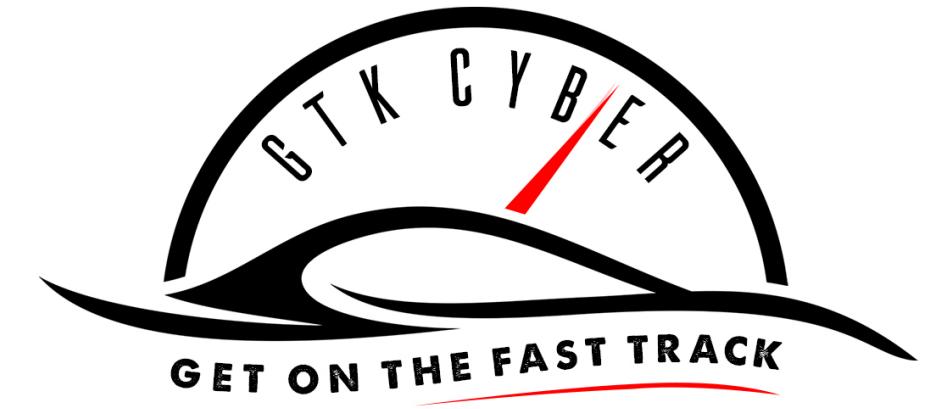




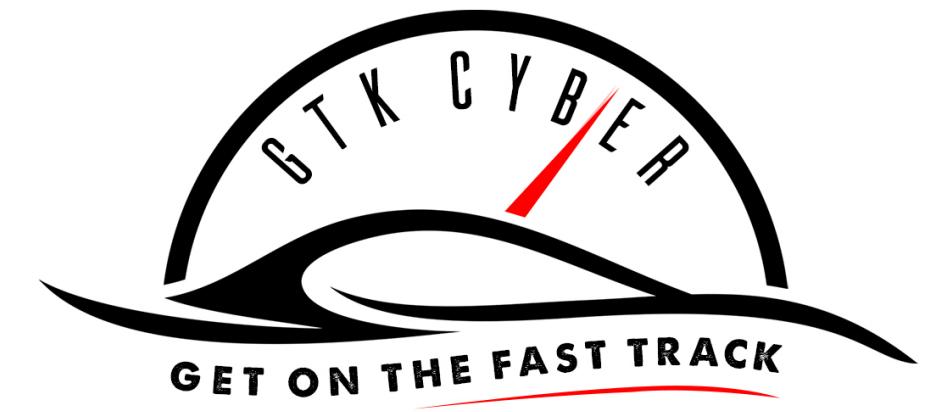
First, define your analytic
question.



What are you trying to do?

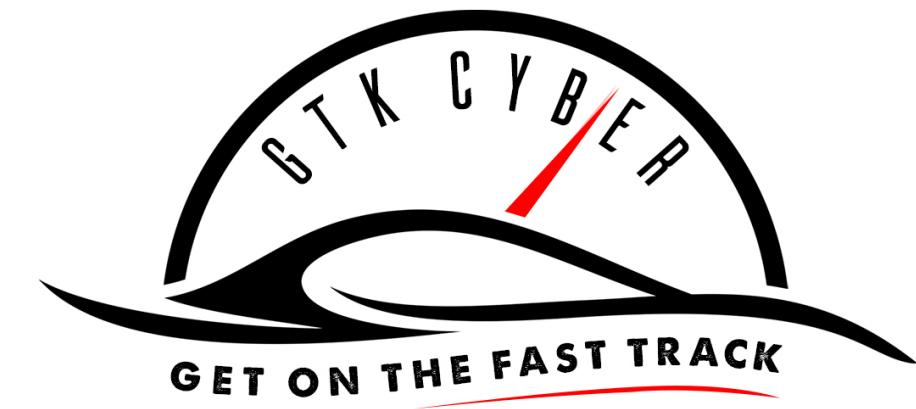


How do you define success?
What are you measuring?



Choose data sources

- What is available?
- Is it enough?
- Is the data reliable/clean/consistent?
- What other data could you use?

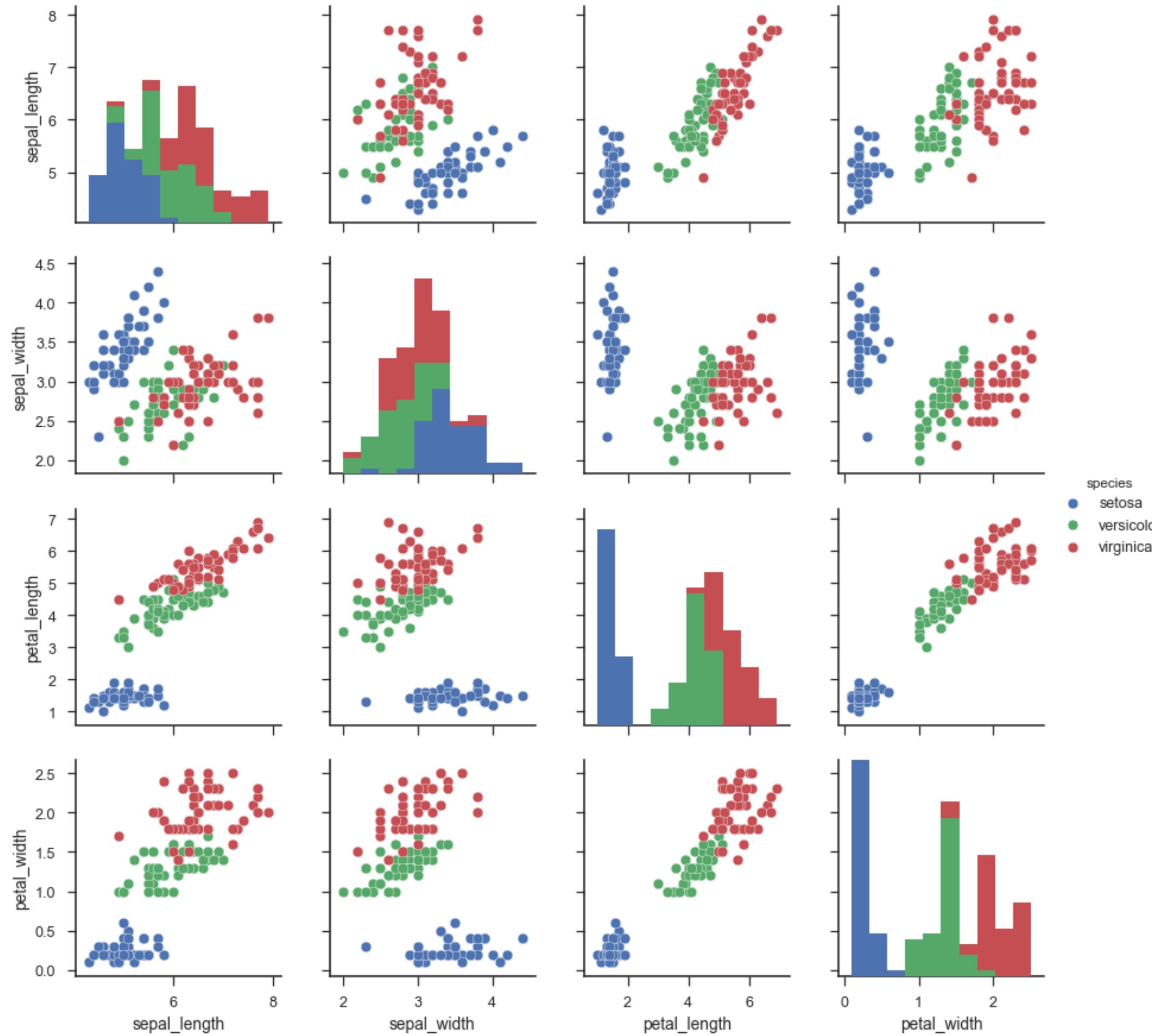


Other Considerations

- Policies
- Legal constraints
- Biases in Data
- Latency
- Data size



Gather and Explore Your Data

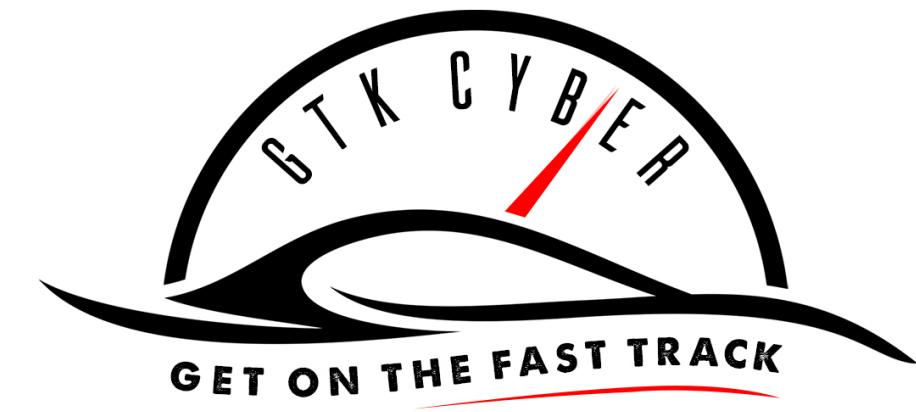


Is the data good enough?

What are the rules governing its use?

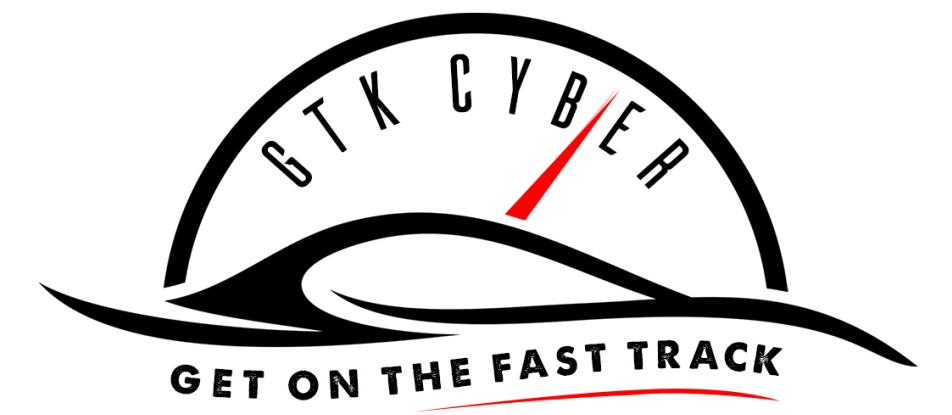
Do I have enough?

Do problems or biases exist in the data that could cause problems?

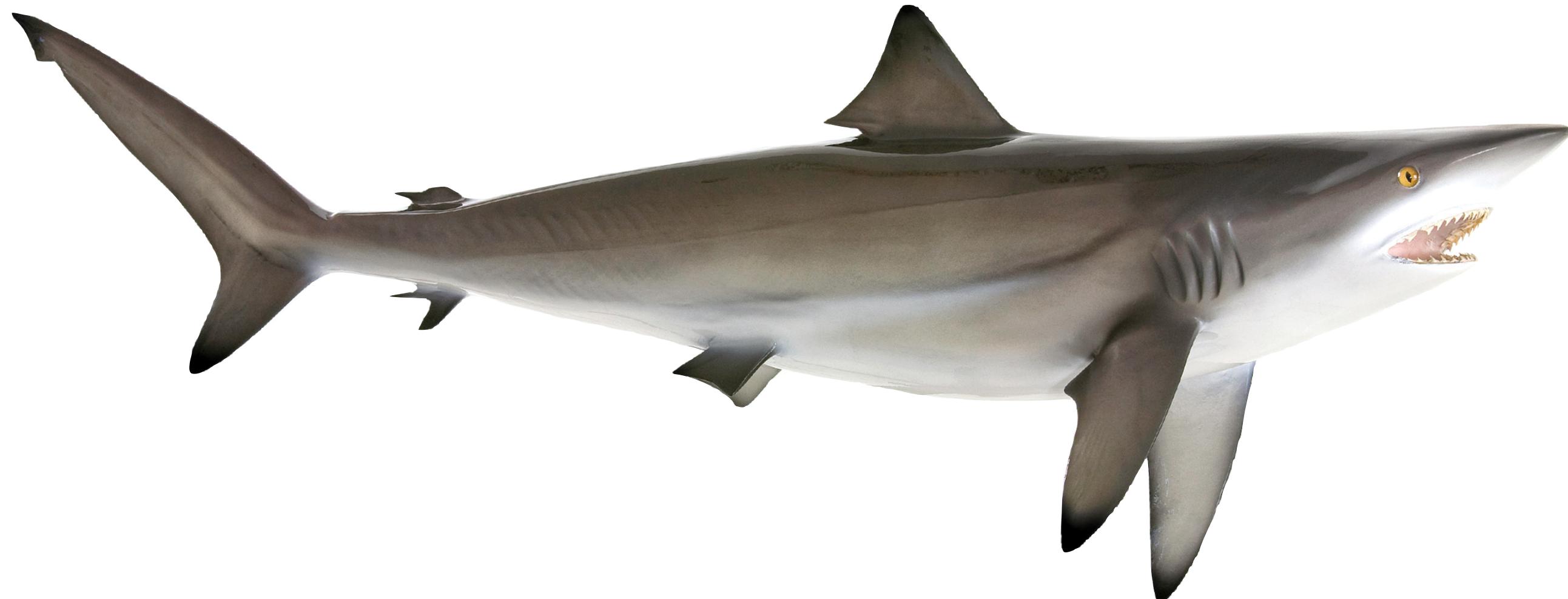


Feature Engineering

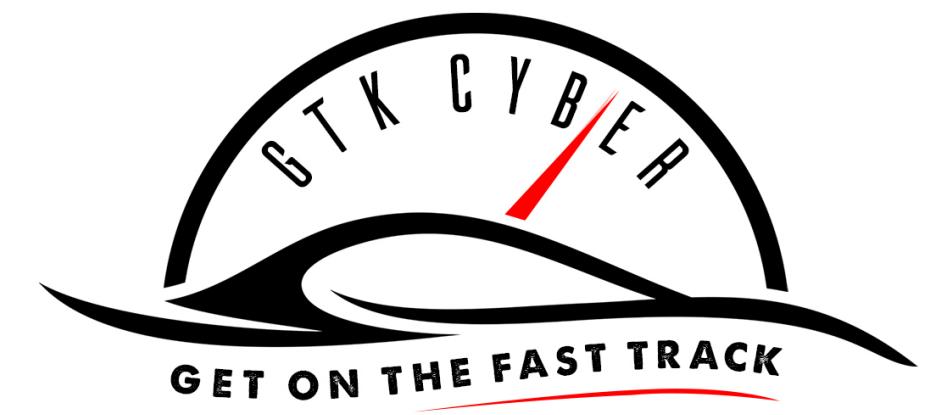
- Define what you are trying to measure. These will become the observations or rows of your final dataset
- Define how you will mathematically represent your data. This will become the features or columns of your final dataset.



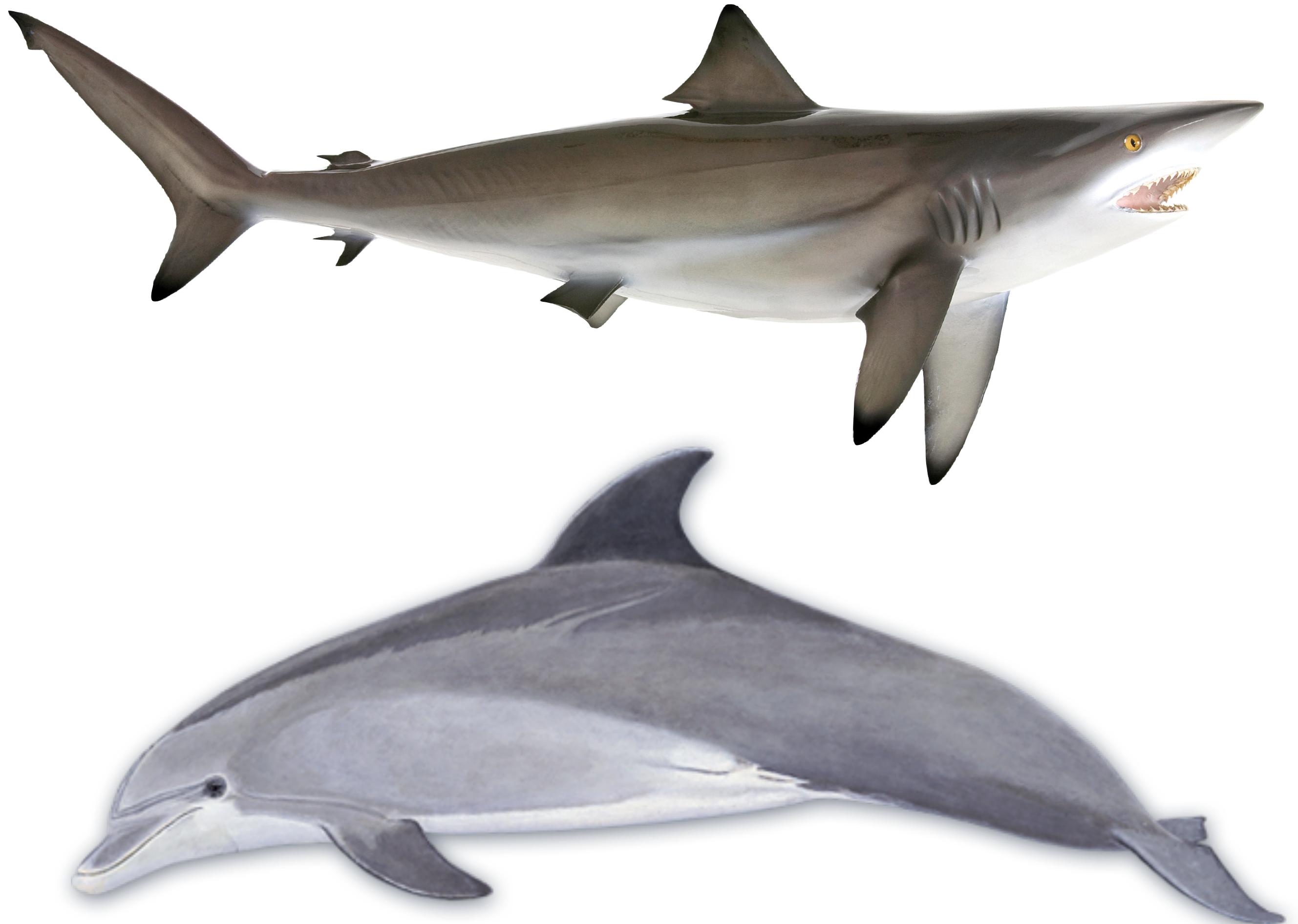
Feature Engineering



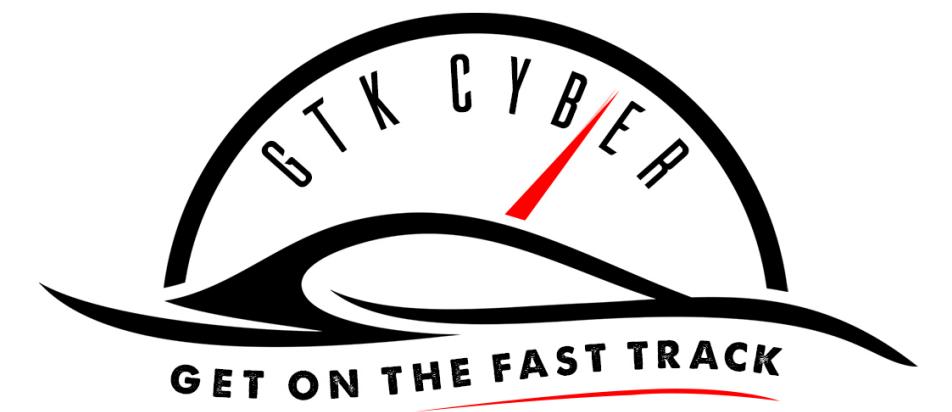
Feature	Value
Color	Gray
Fins	7
Predator	TRUE



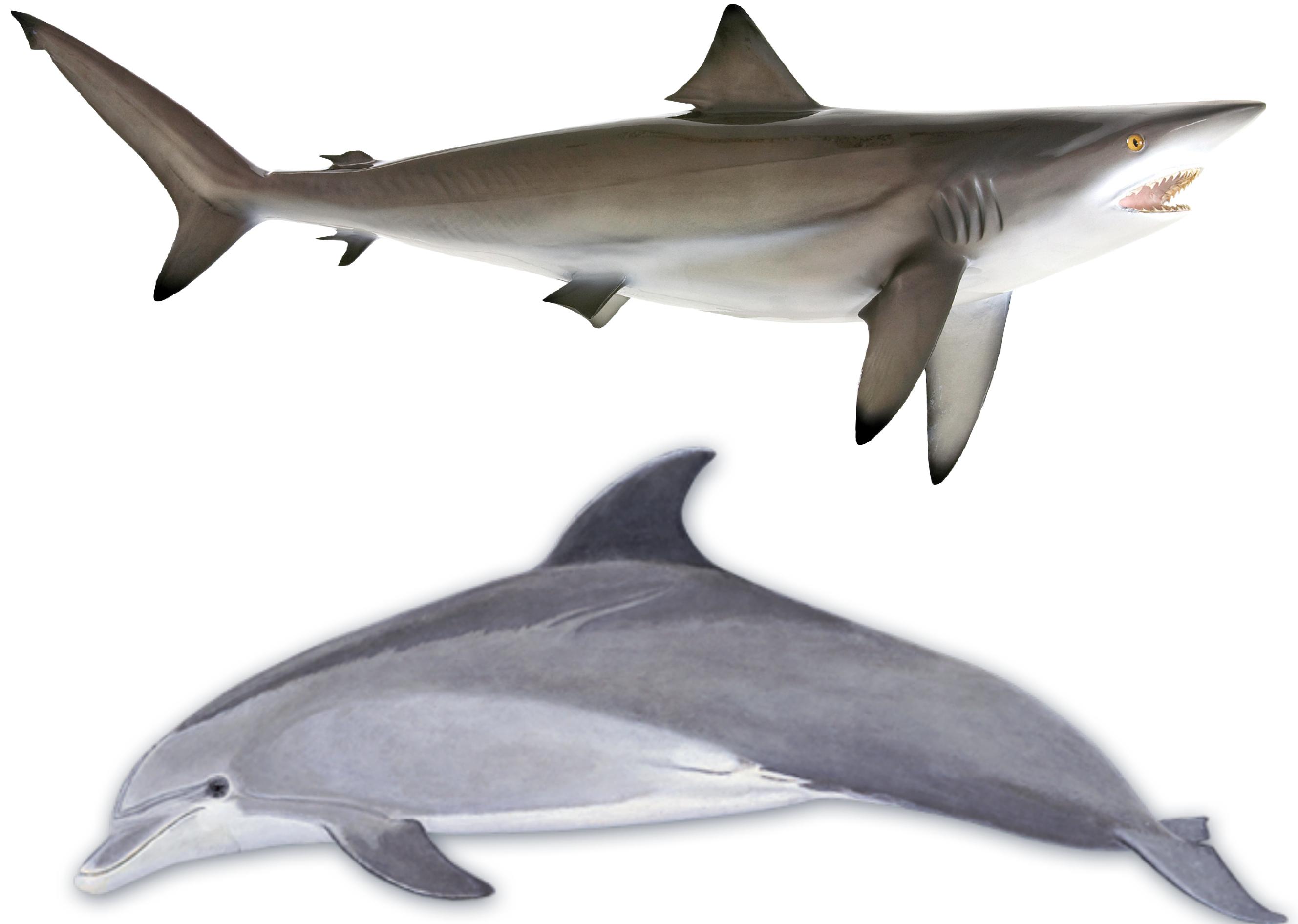
Feature Engineering



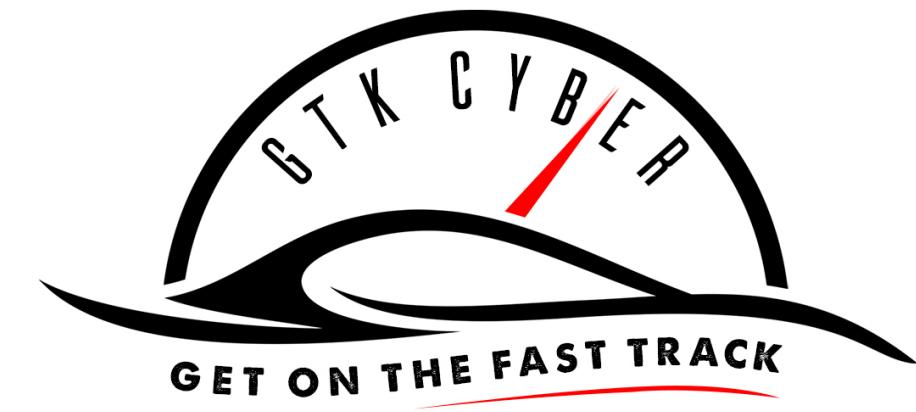
Feature	Value
Color	Gray
Fins	7
Predator	TRUE



Feature Engineering

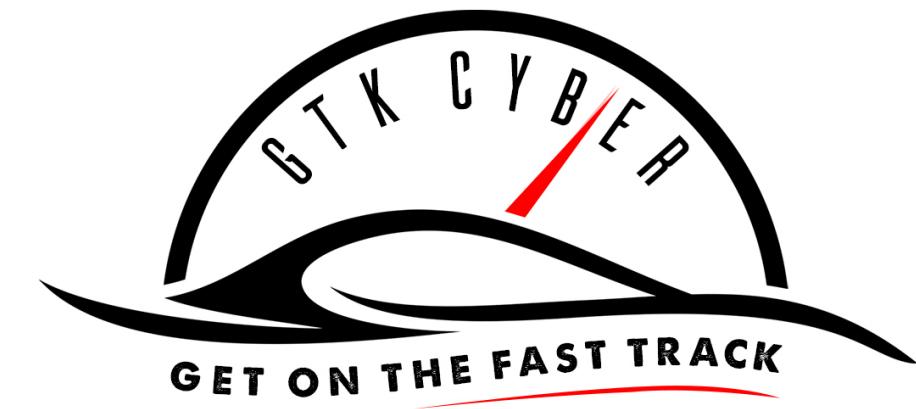


Feature	Value
Color	Gray
Fins	7
Predator	TRUE
Mammal	TRUE



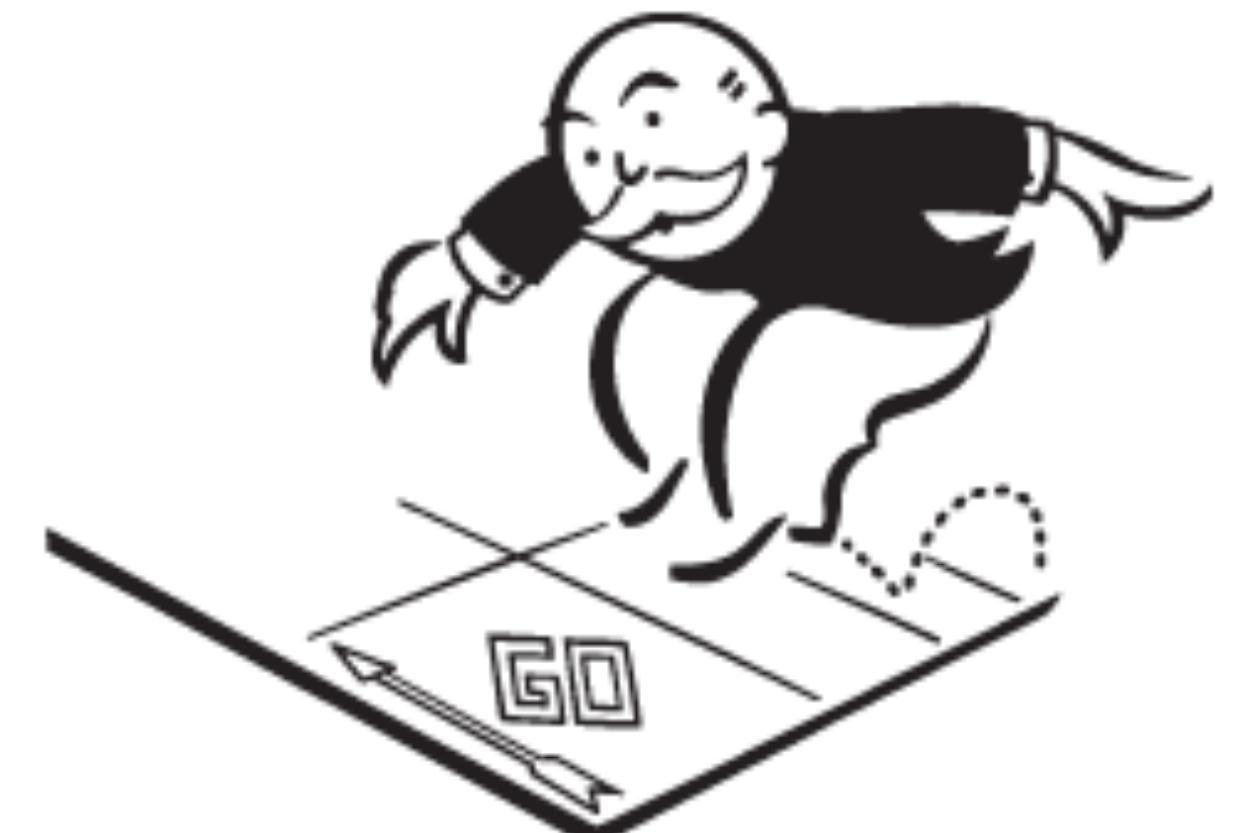
Build and Tune your Model

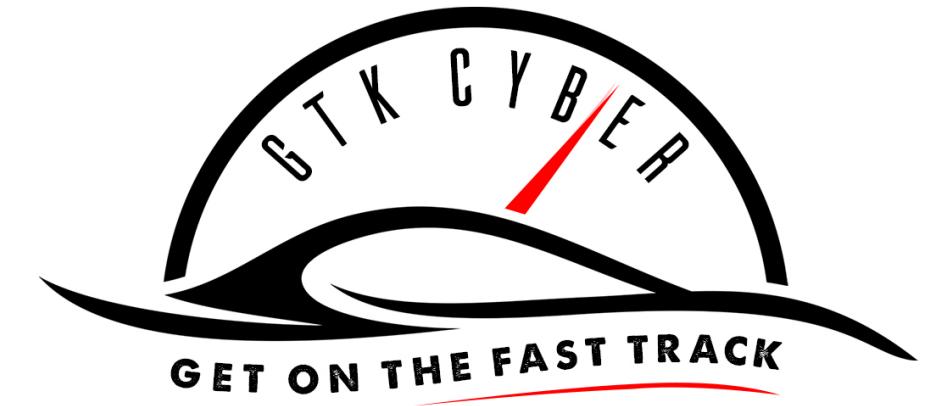
- Believe it or not, this is the easy part.
- Most of this is **done using libraries** like scikit-learn, mllib, tensorflow, caret or keras, and **many steps can be automated**.
- You can even do it in Splunk.



Evaluate Performance

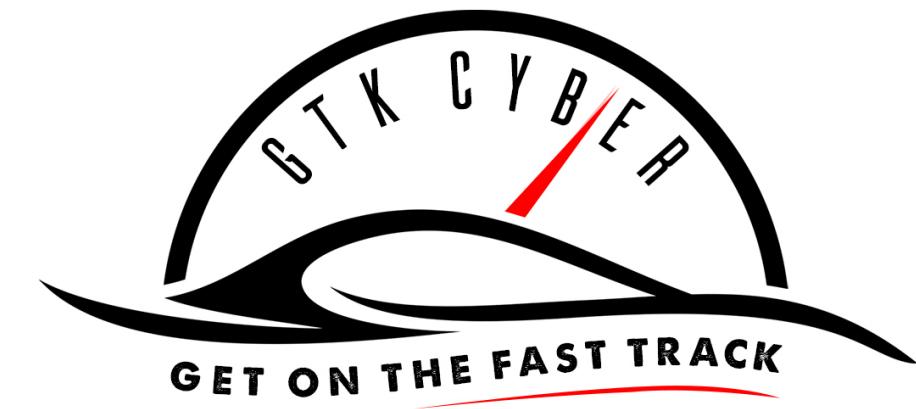
- Use various scoring methods, or write your own to determine model performance.
- Go back to step 1 and repeat! (Do not pass go, do not collect \$200)



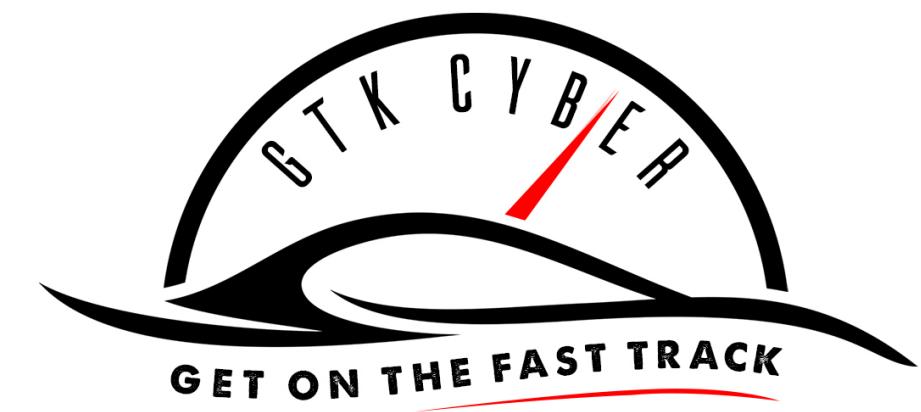


Discussion

- Consider that you are building a system to identify fraudulent credit card transactions. What are some features that you would want to capture?
- What are the challenges you could foresee in this problem?

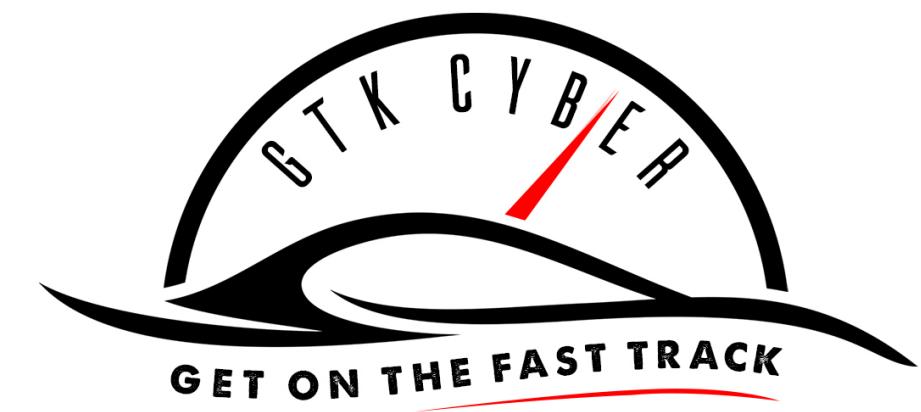


The Python Data Science Ecosystem

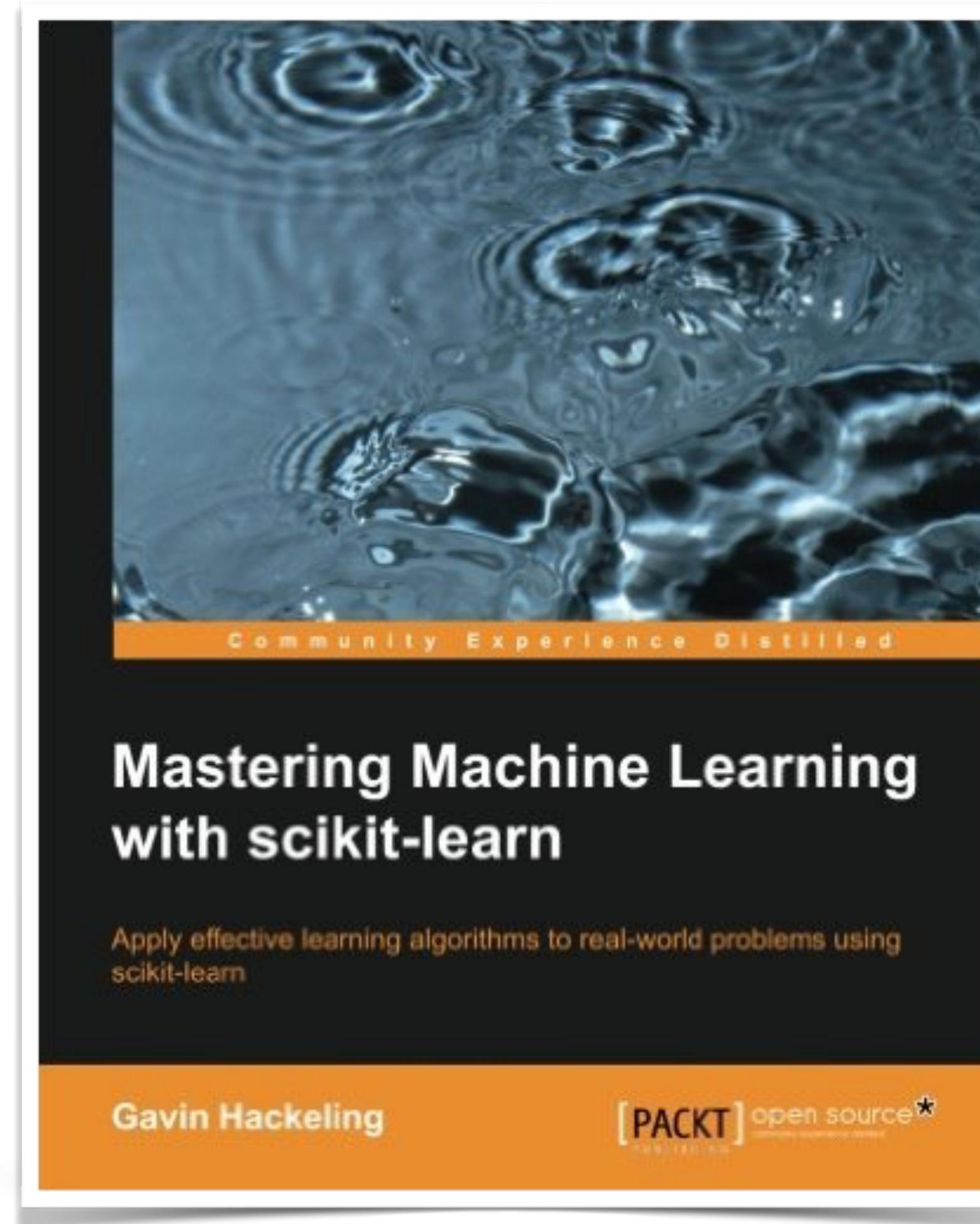


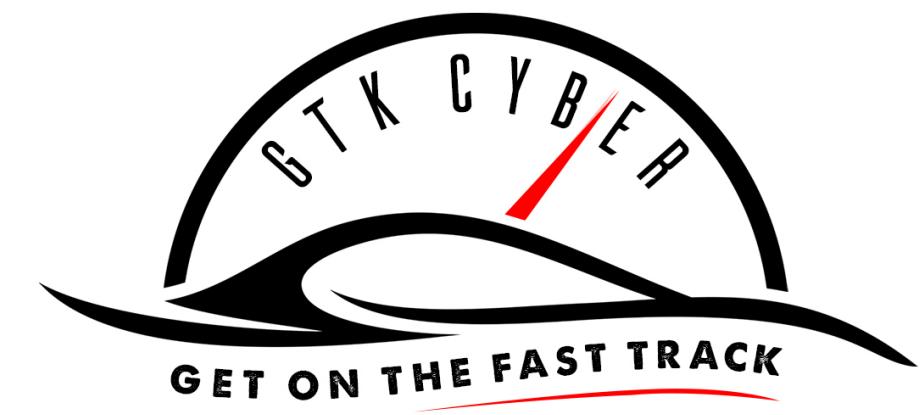
Machine Learning Ecosystem

- **Data Gathering:** Pandas, Drill, BeautifulSoup, PyDBAPI, PyDAL, Boto3
- **Feature Extraction:** Pandas, NumPy, Featuretools
- **Machine Learning**
 - **"Regular" ML:** Scikit-learn (sklearn), h2o, mllib (PySpark)
 - **Deep Learning:** Tensorflow, Keras, Theano, Caffe, PyTorch
- **Visualization:** Matplotlib, Seaborn, Yellowbrick, LIME, ggplot, plot.ly,

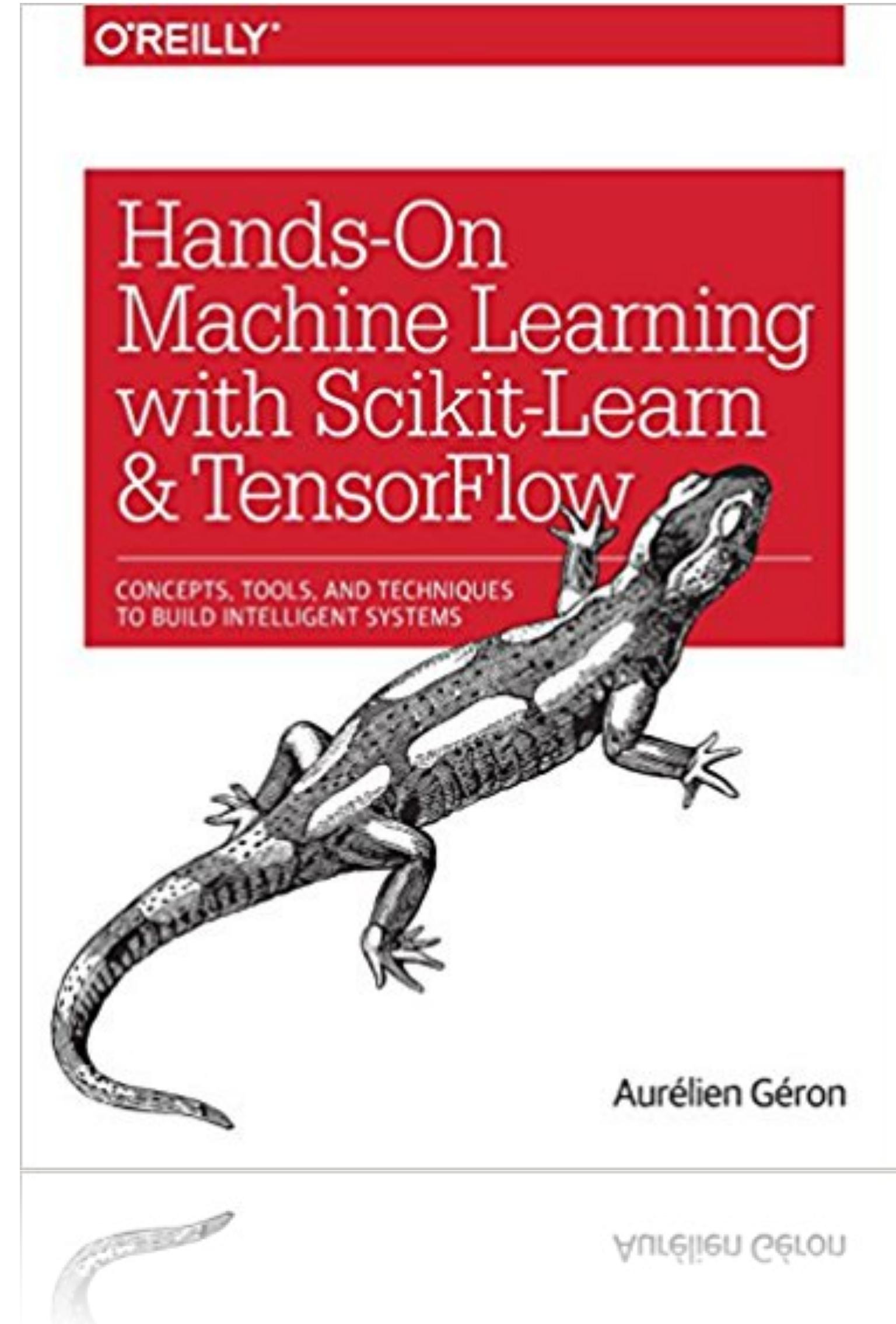


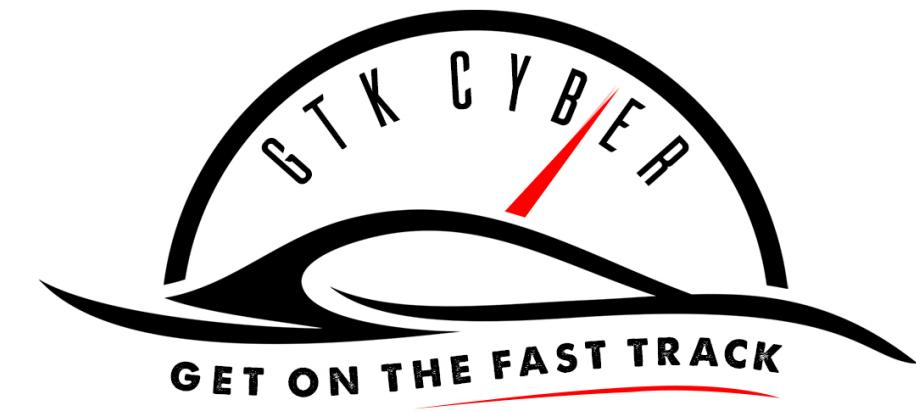
Recommended Reading



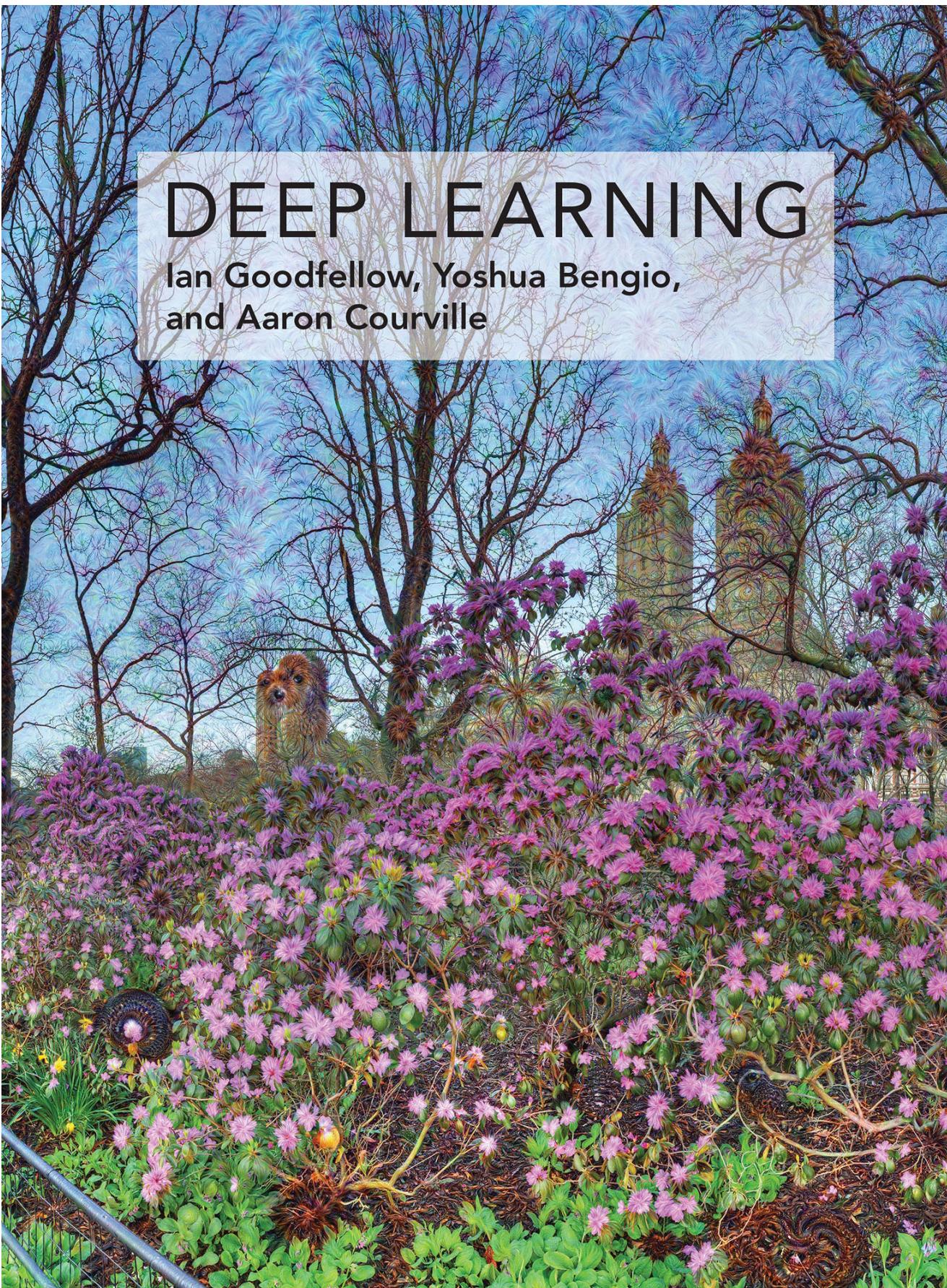


Recommended Reading

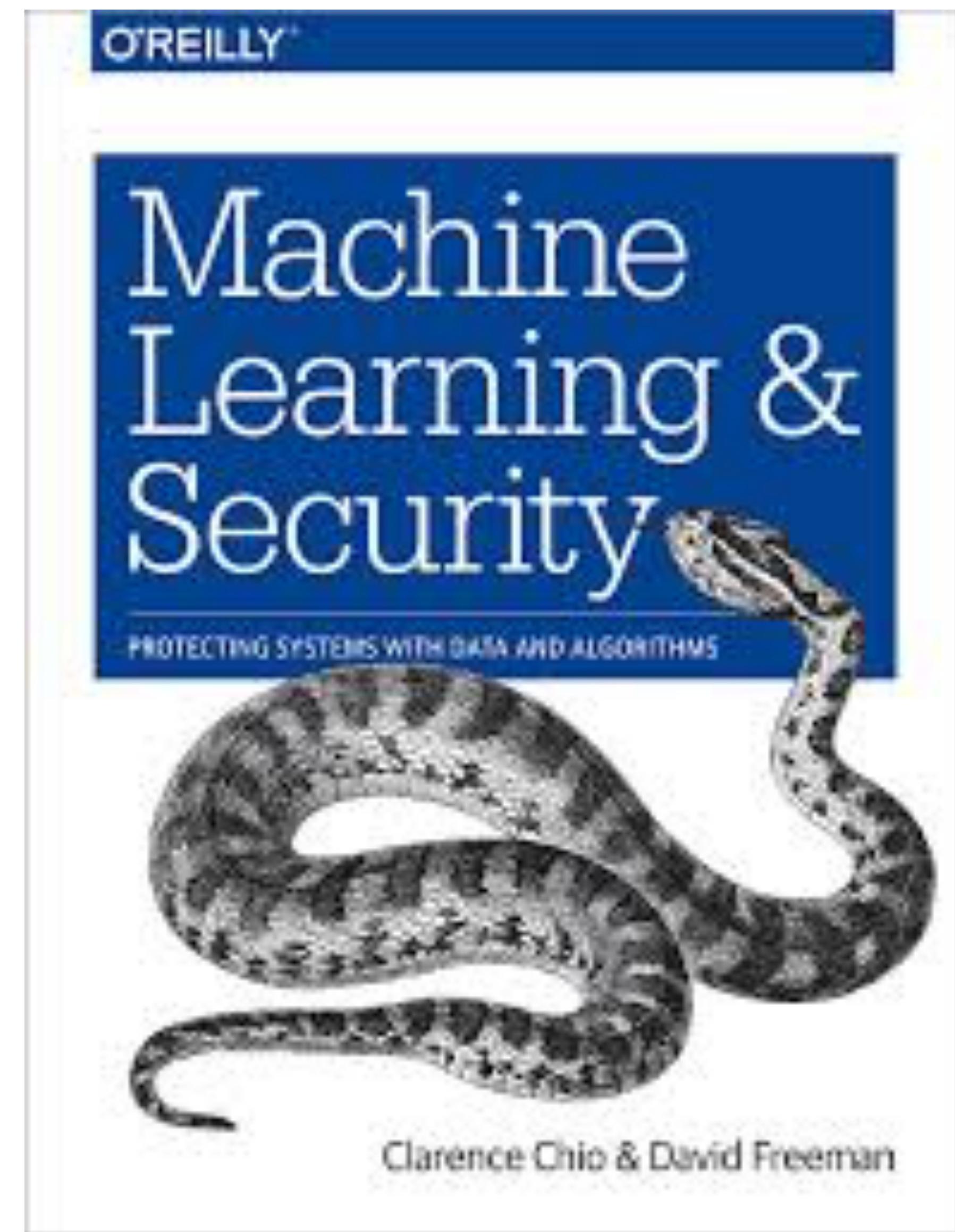
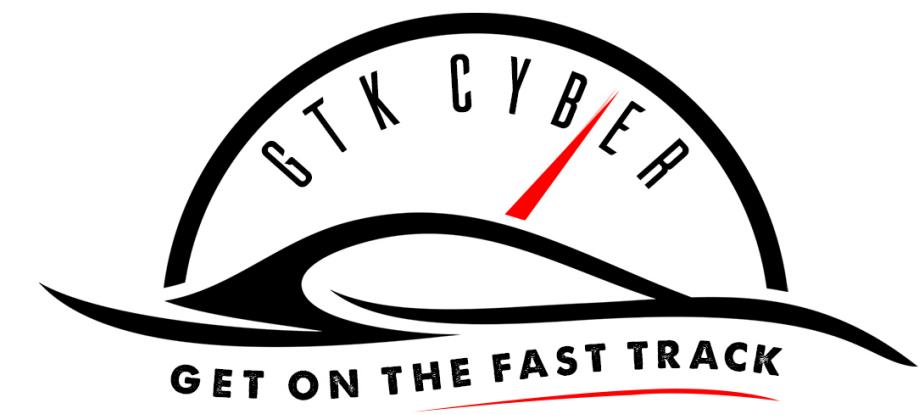


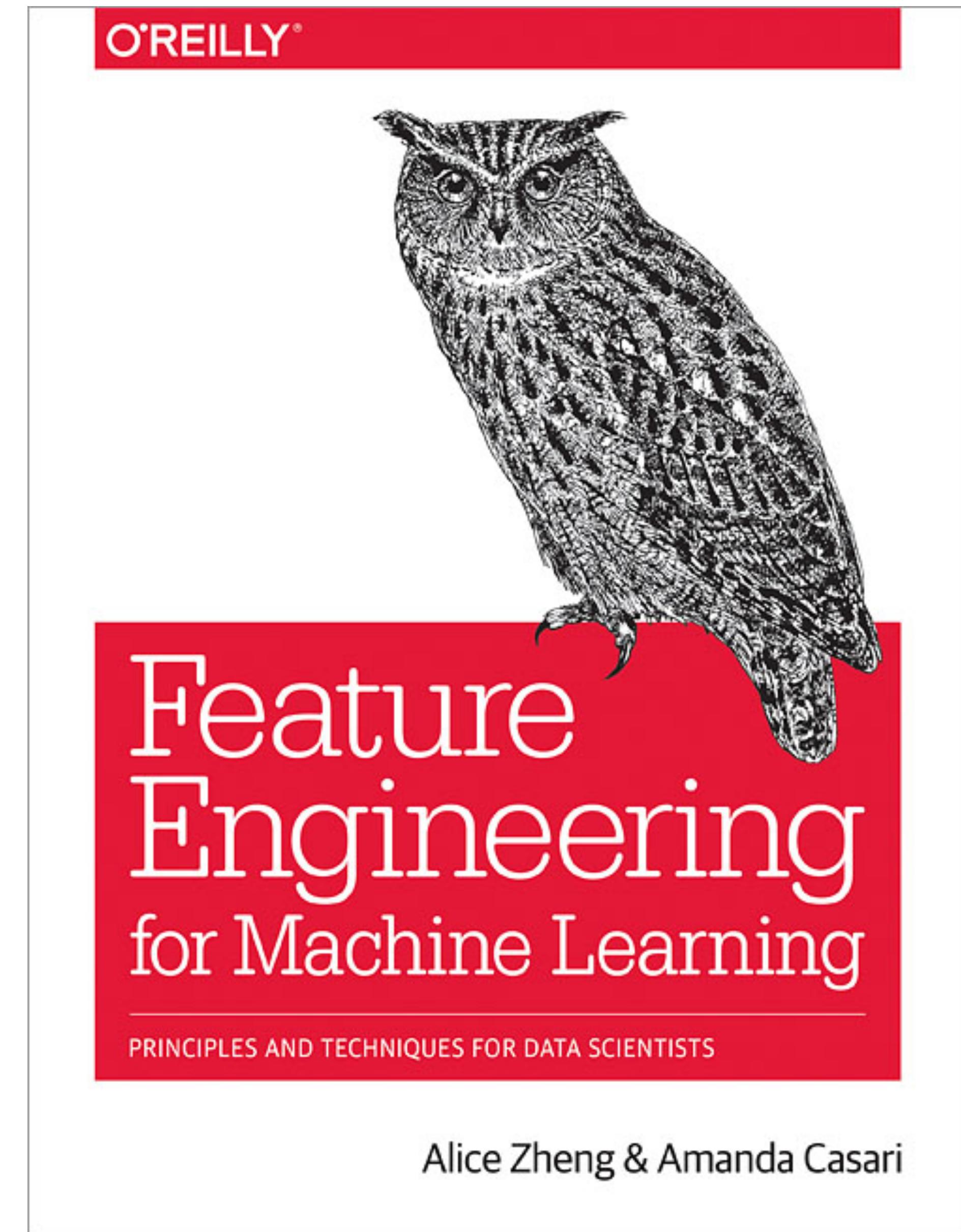
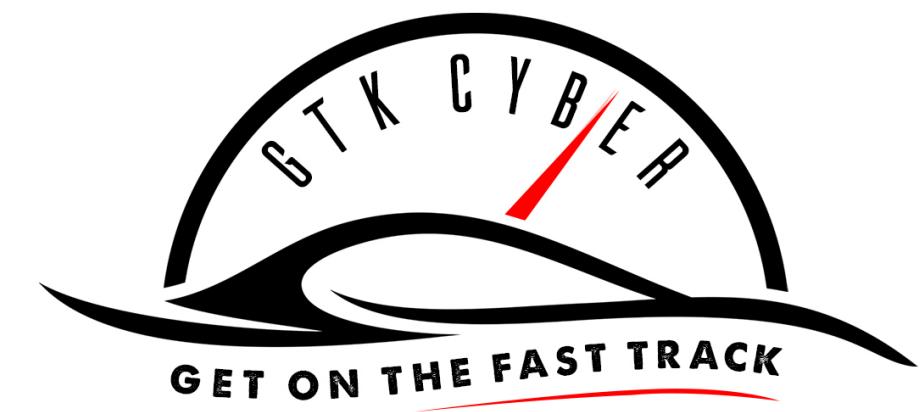


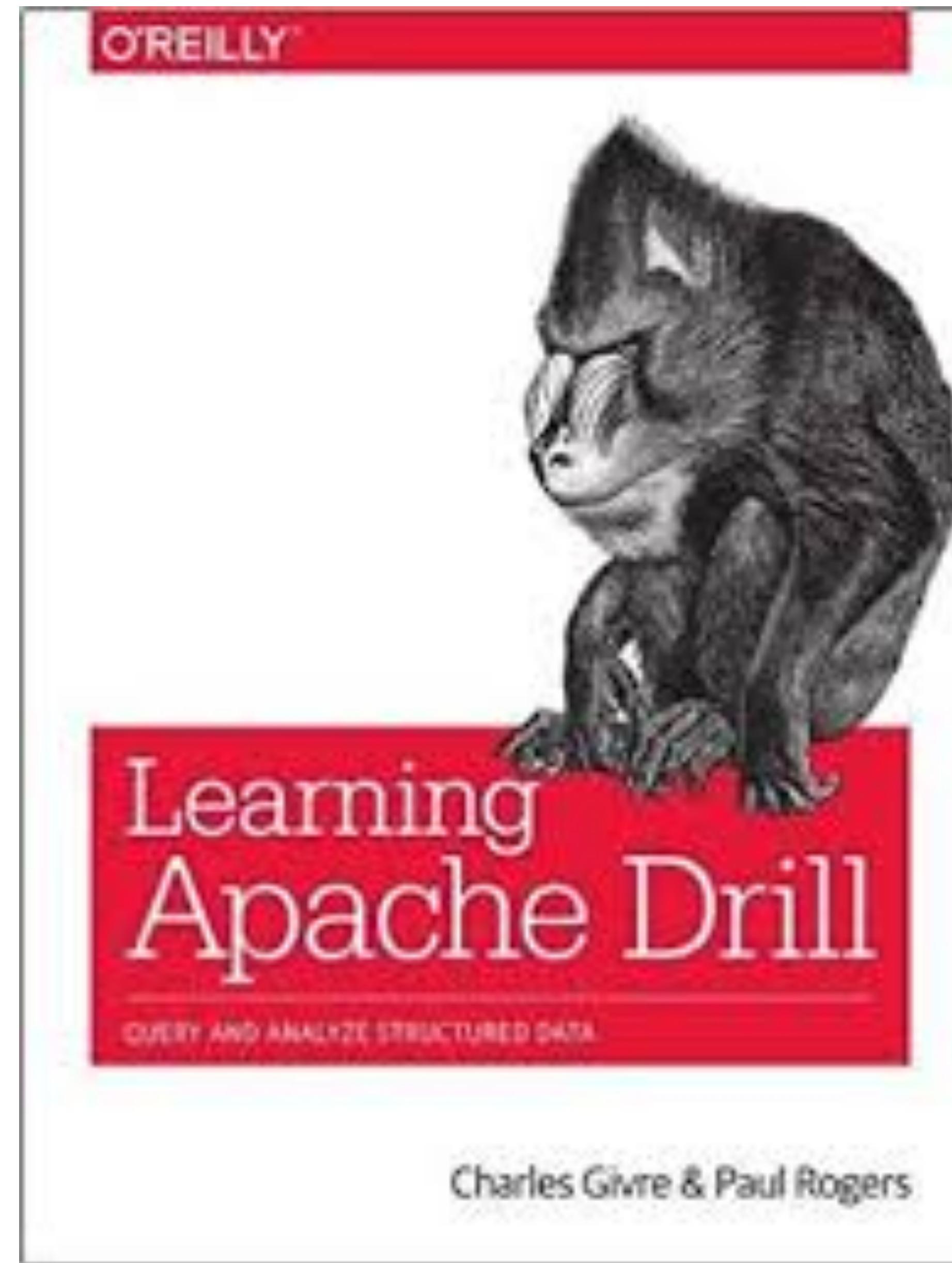
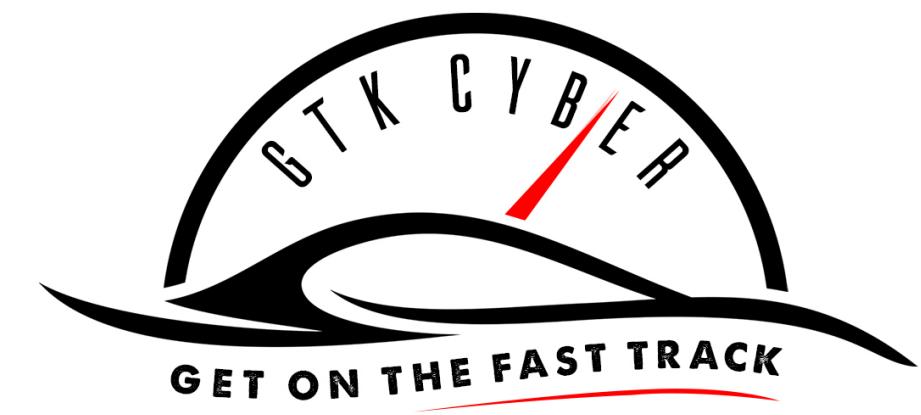
Recommended Reading

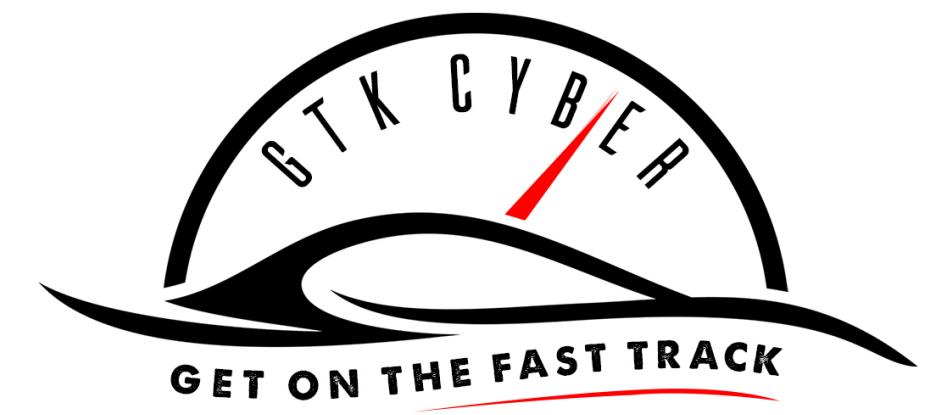


<http://www.deeplearningbook.org/>





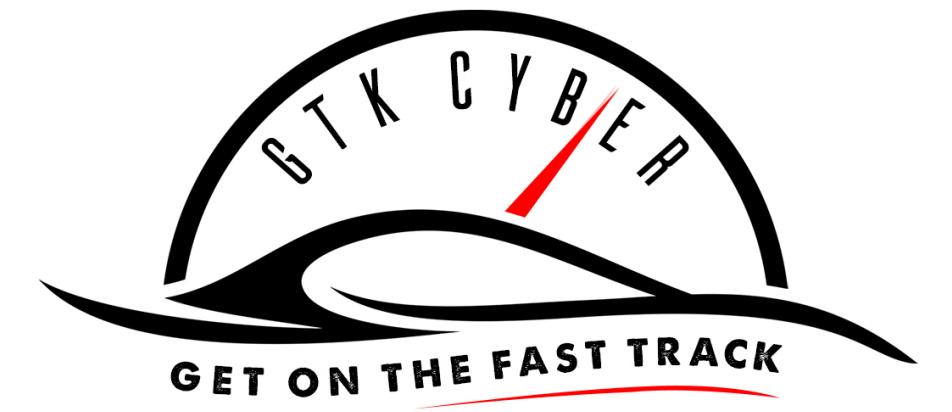




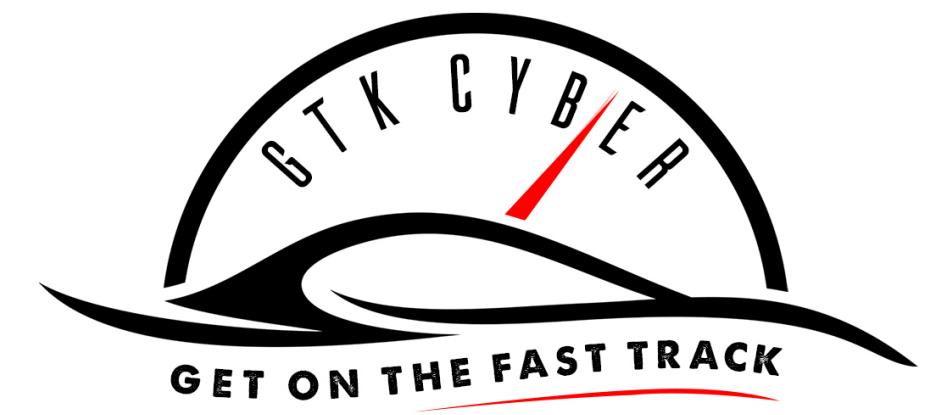
The Virtual Machine: Griffon



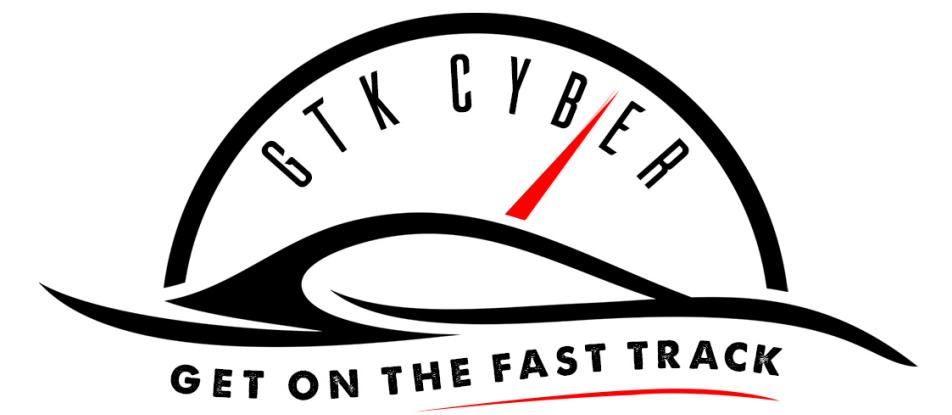
```
File Edit View Search Terminal Help
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hbase/hbase-1.1.3/lib/slf4j-log4j12-1
.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4
j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2016-05-16 13:04:54,887 WARN  [main] util.NativeCodeLoader: Unable to load nativ
e-hadoop library for your platform... using builtin-jar classes where applicabl
e
2016-05-16 13:05:11,827 ERROR [main] zookeeper.RecoverableZooKeeper: ZooKeeper exists faile
d after 4 attempts
2016-05-16 13:05:11,828 WARN  [main] zookeeper.ZKUtil: hconnection-0x46a145ba0x0, quorum=lo
calhost:2181, baseZNode=/hbase Unable to set watcher on znode (/hbase/hbaseid)
org.apache.zookeeper.KeeperException$ConnectionLossException: KeeperErrorCode = ConnectionL
oss for /hbase/hbaseid
        at org.apache.zookeeper.KeeperException.create(KeeperException.java:99)
        at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
        at org.apache.zookeeper.ZooKeeper.exists(ZooKeeper.java:1045)
        at org.apache.hadoop.hbase.zookeeper.RecoverableZooKeeper.exists(RecoverableZooKeep
er.java:221)
        at org.apache.hadoop.hbase.zookeeper.ZKUtil.checkExists(ZKUtil.java:541)
        at org.apache.hadoop.hbase.zookeeper.ZKClusterId.readClusterIdZNode(ZKClusterId.jav
a:65)
        at org.apache.hadoop.hbase.client.ZooKeeperRegistry.getClusterId(ZooKeeperRegistry.
java:105)
```



Do Data Science, Not Sysadmin



Built on Ubuntu MATE



Easy to Use

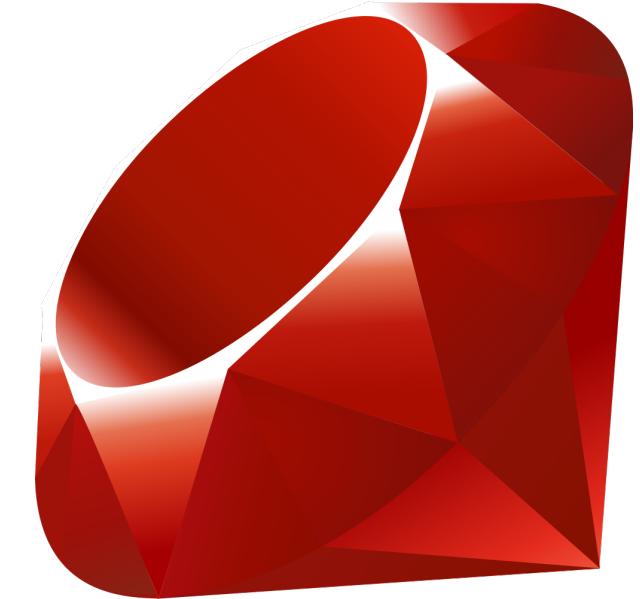
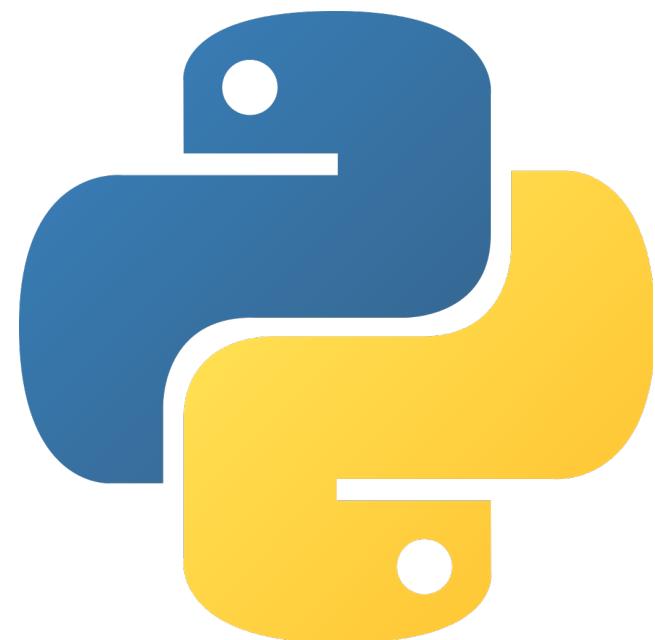


Programming Languages

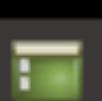
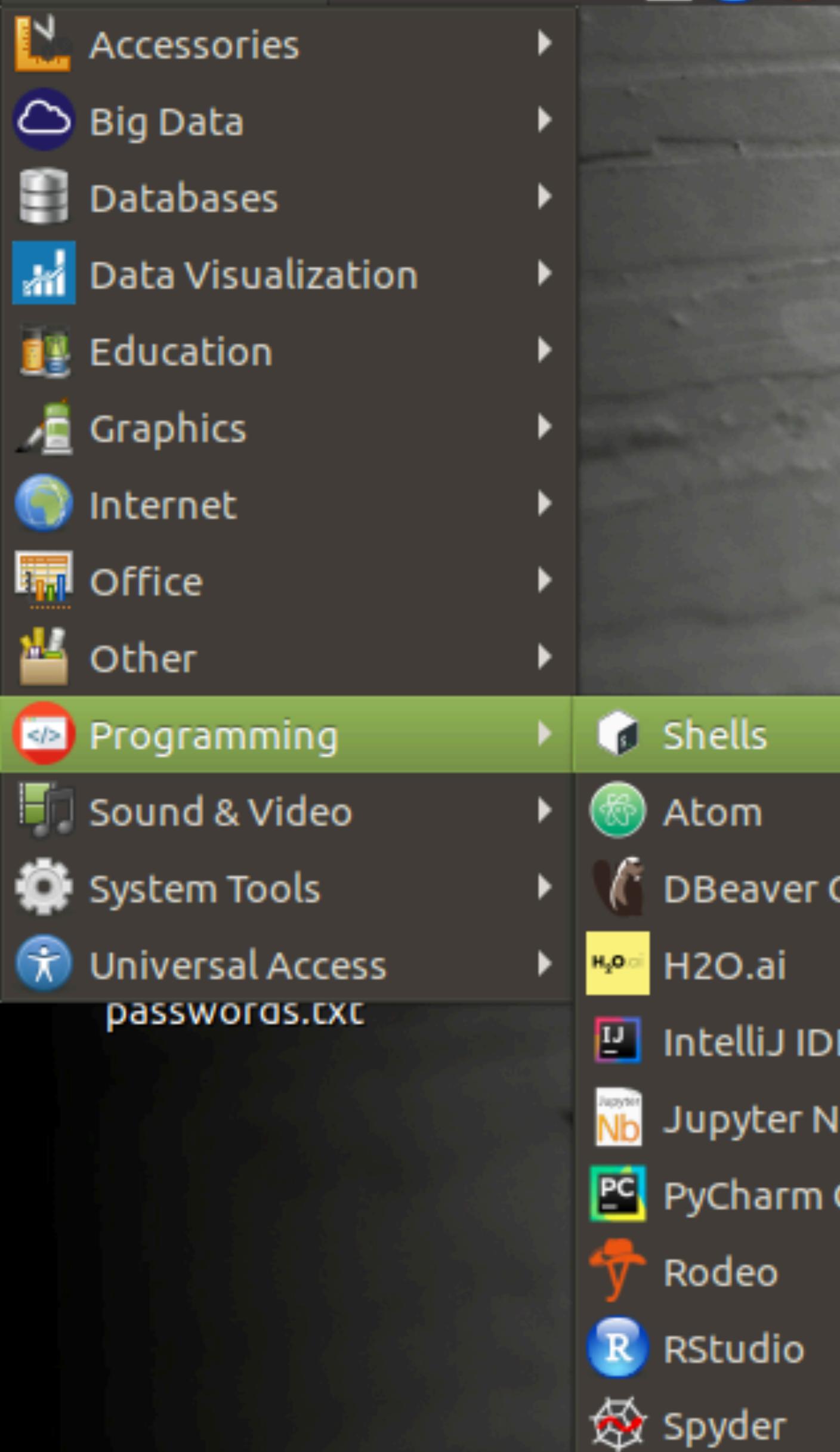
- Languages
- Libraries
- Editors and Notebooks
- Databases + Administrative tools
- Big Data Tools
- Machine Learning Libraries
- Data Visualization

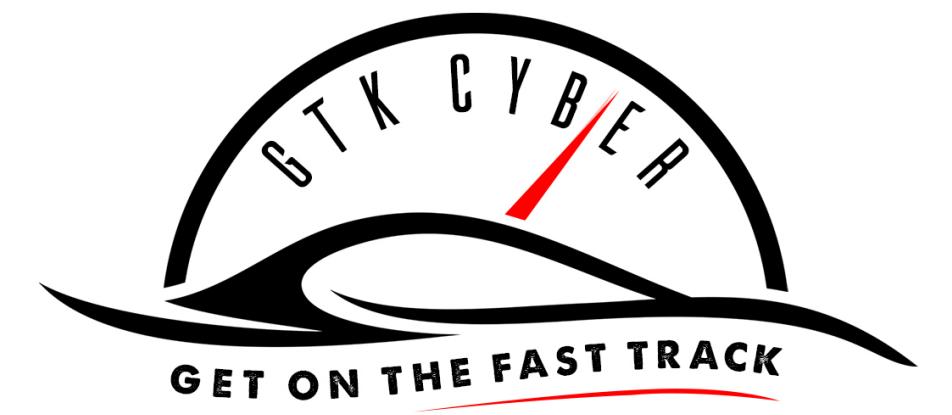


Scripting Languages

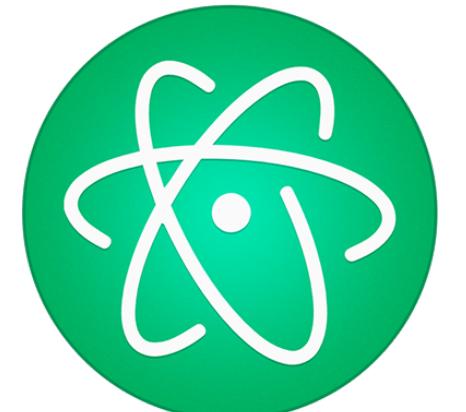
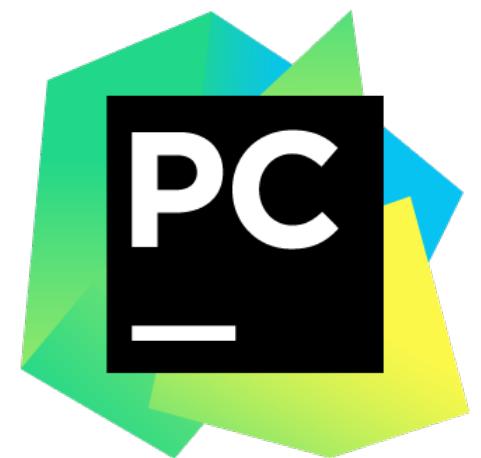
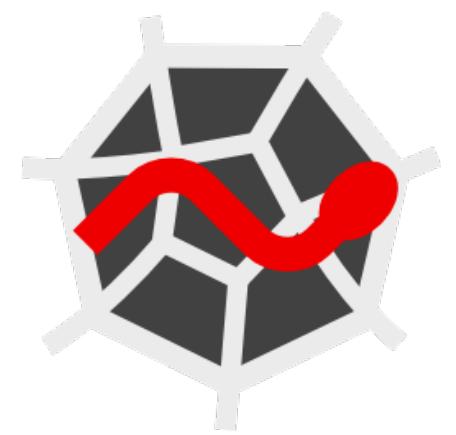
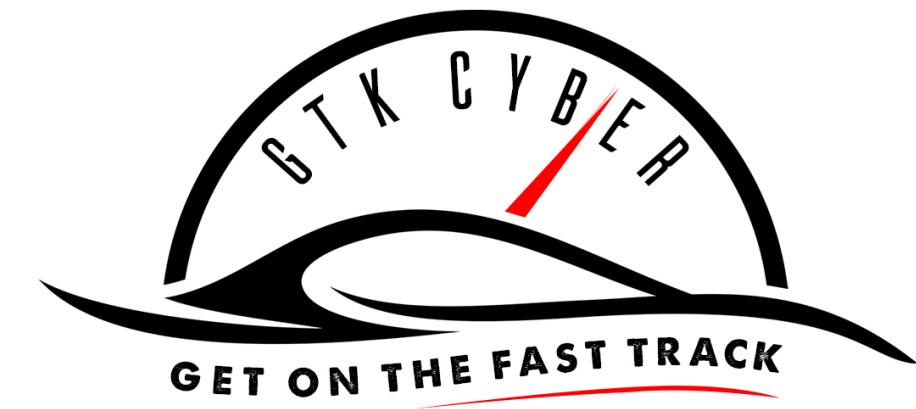


Applications Places System



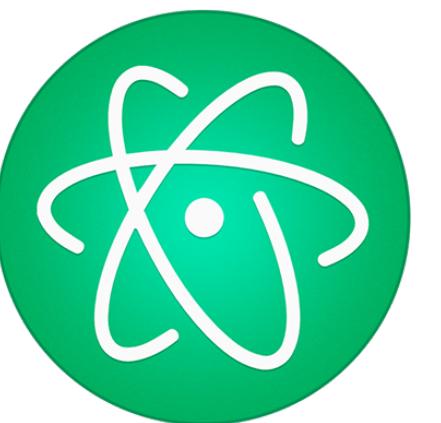
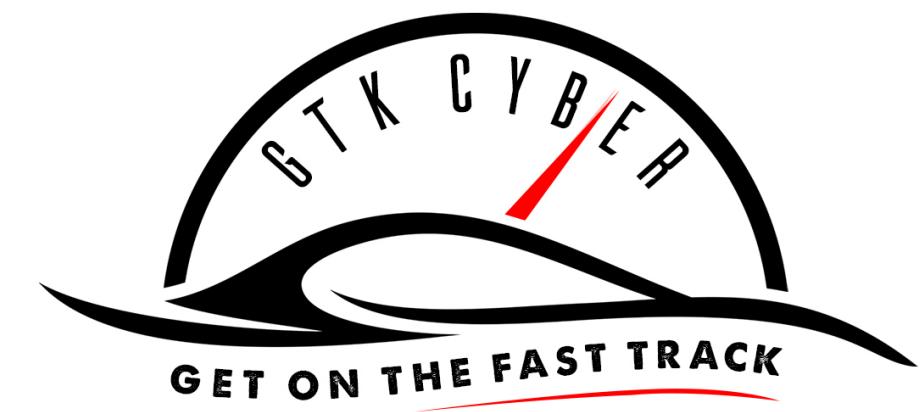


Editors & Notebooks



gtkcyber.com





Atom

Project Merlin Beta 3 (Small updates) [Running]

Applications Places System R Jupyter Nb Mon Aug 15, 12:26

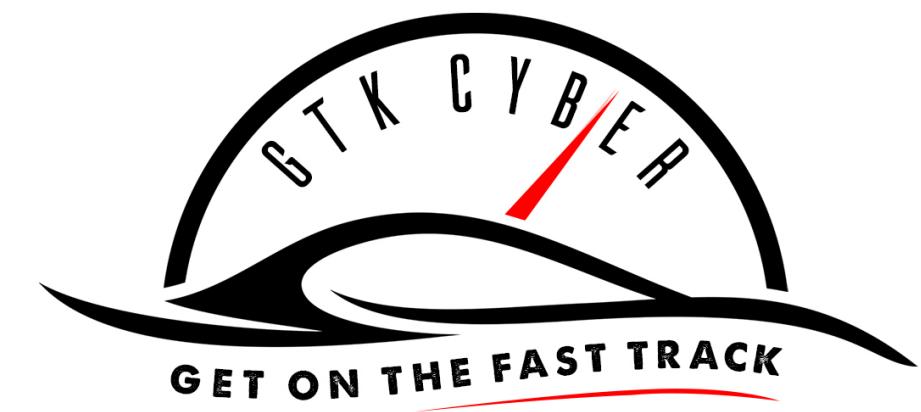
python_demo.py — /usr/share/atom/resources — Atom

File Edit View Selection Find Packages Help

resources

```
python_demo.py
```

```
1 import pandas as pd
2 df = pd.DataFrame([2,3,4,5,6,7,8])
```



gtkcyber.com

The screenshot shows a Jupyter Notebook interface running in a Mozilla Firefox browser window titled "Project Merlin Development Version (Alpha 0.2) [Running]". The browser's toolbar includes icons for Applications, Places, System, Firefox, R, and Nb. The main content area displays the Jupyter logo and a file tree under the "Files" tab. The file tree lists several directories: anaconda3, Desktop, Documents, Downloads, metastore_db, Music, and node_modules. To the right of the file tree is a sidebar with a "New" dropdown menu containing options for Text File, Folder, Terminal, Notebooks, Julia 0.4.2, Python 2, Python 3, R, Ruby 2.1.5, Scala 2.11, and pySpark (Spark 1.6.0). The status bar at the bottom shows the date and time: "Tue May 10, 23:47".

Project Merlin Development Version (Alpha 0.2) [Running]

Applications Places System Firefox R Nb

Tue May 10, 23:47

File Edit View History Bookmarks Tools Help

Home - Mozilla Firefox

Home

localhost:8888/tree

Search

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New

Text File
Folder
Terminal

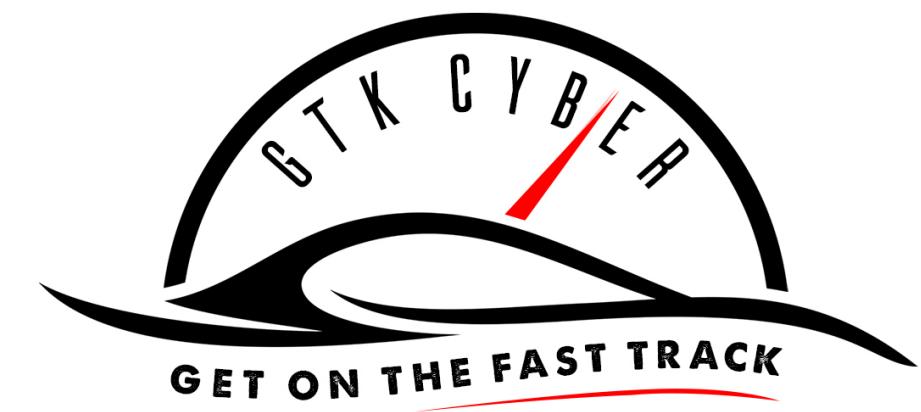
Notebooks
Julia 0.4.2
Python 2
Python 3
R
Ruby 2.1.5
Scala 2.11
pySpark (Spark 1.6.0)

anaconda3
Desktop
Documents
Downloads
metastore_db
Music
node_modules



Jupyter Notebook

- Python 2 & 3
- R
- Ruby
- Scala
- PySpark

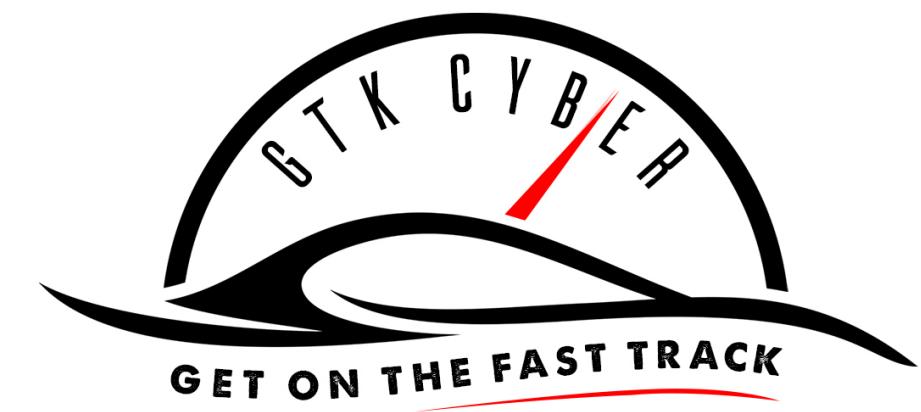


gtkcyber.com



mongoDB®





MySQL®

Project Merlin Beta 3 (Before beaker 1.6 update) [Running]

Fri Aug 19, 00:14

MySQL Workbench

Local instance 3306 Local instance 3306

File Edit View Query Database Server Tools Scripting Help

SQL

MANAGEMENT

- Server Status
- Client Connections
- Users and Privileges
- Status and System Variables
- Data Export
- Data Import/Restore

INSTANCE

- Startup / Shutdown
- Server Logs
- Options File

PERFORMANCE

- Dashboard
- Performance Reports
- Performance Schema Setup

SCHEMAS

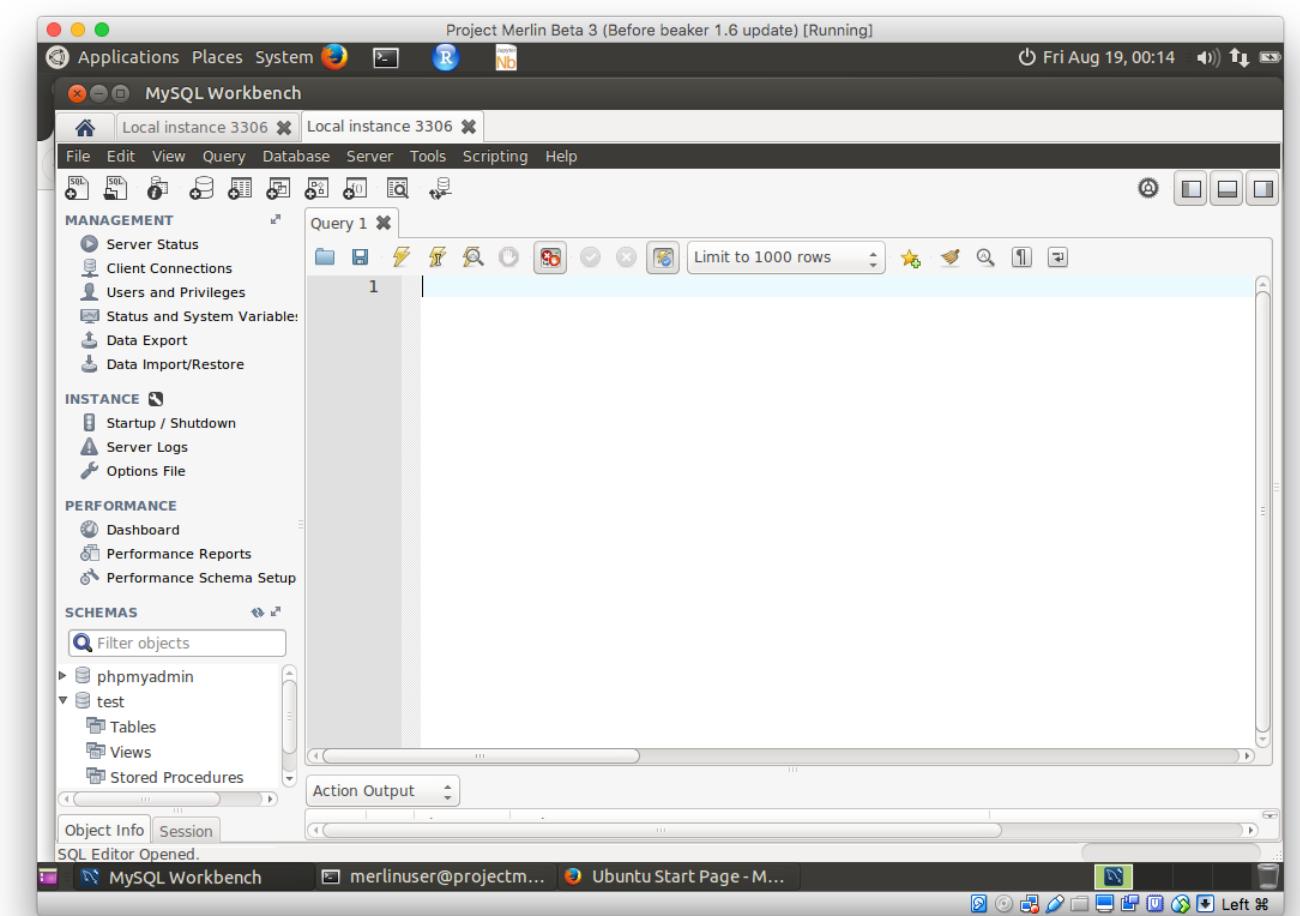
Filter objects

- phpmyadmin
- test
 - Tables

Query 1

Limit to 1000 rows

1



Project Merlin Beta 3 (Before beaker 1.6 update) [Running]

Fri Aug 19, 00:14

localhost / localhost | phpMyAdmin 4.4.13.1deb1 - Mozilla Firefox

PMA localhost / localhost... +

localhost / phpmyadmin/index.php?token=62fb6ffb630fba16f23788ee21dc4e

Search

Applications Places System R Nb

phpMyAdmin

Recent Favorites

Server: localhost

Databases SQL Status Users Export Import Settings More

General Settings

Change password

Server connection collation: utf8mb4_unicode_ci

Database server

- Server: Localhost via UNIX socket
- Server type: MySQL
- Server version: 5.6.31-0ubuntu0.15.10.1 - (Ubuntu)
- Protocol version: 10
- User: merlinuser@localhost
- Server charset: UTF-8 Unicode (utf8)

Appearance Settings

Language: English

Theme: pmahomme

Font size: 82%

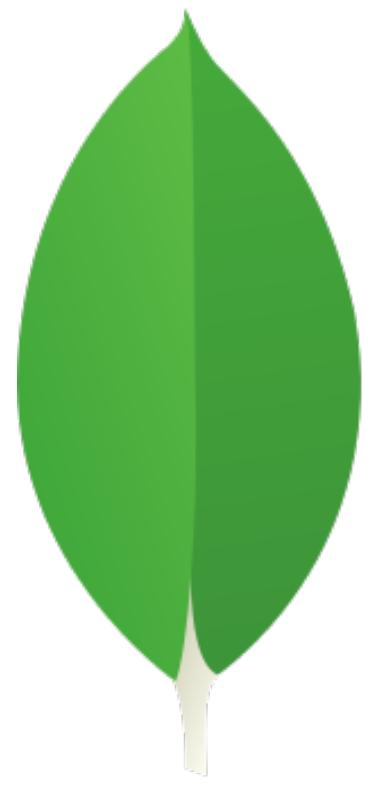
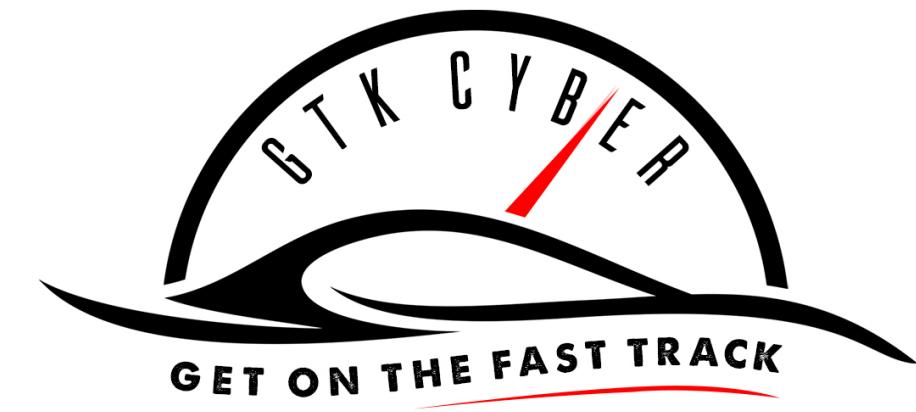
More settings

Web server

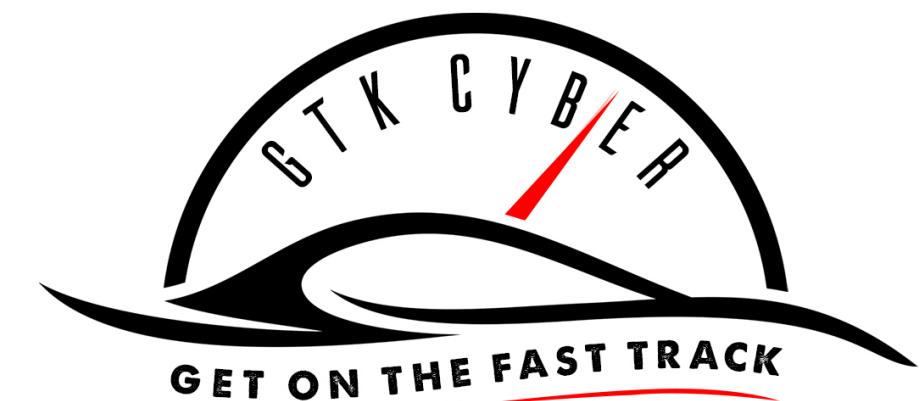
- nginx/1.10.1
- Database client version: libmysql - 5.6.31
- PHP extension: mysqli
- PHP version: 5.6.11-1ubuntu3.4

phpMyAdmin

The screenshot shows the MySQL Workbench interface on the left and the phpMyAdmin interface on the right. The phpMyAdmin interface is running on a local host and displays various configuration settings. The 'General Settings' section includes a 'Change password' link and a dropdown for 'Server connection collation' set to 'utf8mb4_unicode_ci'. The 'Database server' section provides detailed information about the MySQL server setup. The 'Appearance Settings' section allows users to change the language (set to English), theme (set to pmahomme), and font size (set to 82%). The 'Web server' section lists the installed web server (nginx/1.10.1), database client version (libmysql - 5.6.31), PHP extension (mysqli), and PHP version (5.6.11-1ubuntu3.4).



mongoDB®



A screenshot of a Mozilla Firefox browser window. The title bar shows "db - Mongo Express - Mozilla Firefox". The address bar displays "localhost:8081/db/db". The main content area shows the MongoDB Express interface with the heading "Mongo Express Database: db".

Viewing Database: db

Collection "test" deleted!

Collections

Collection Name

+ Create collection

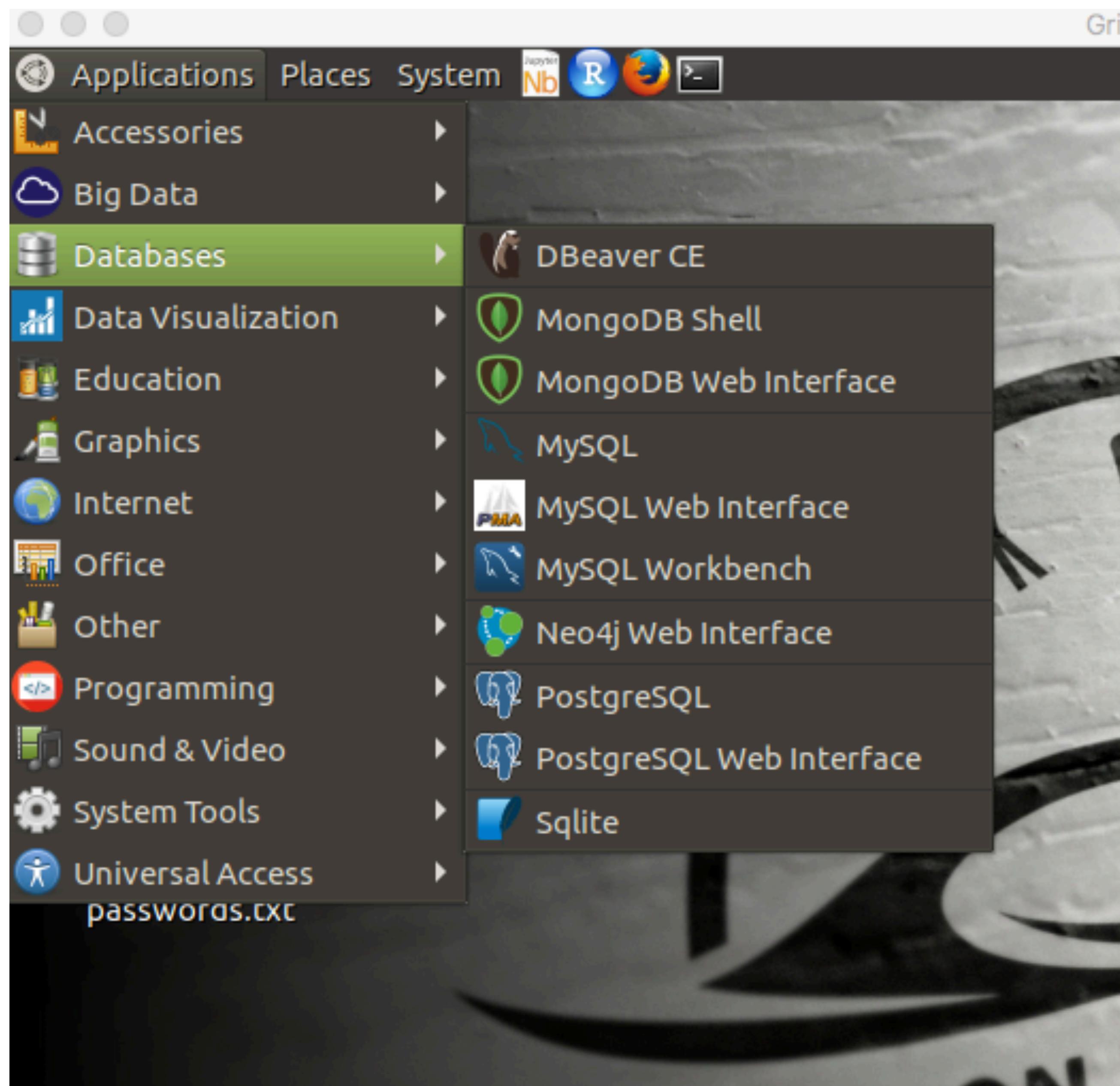
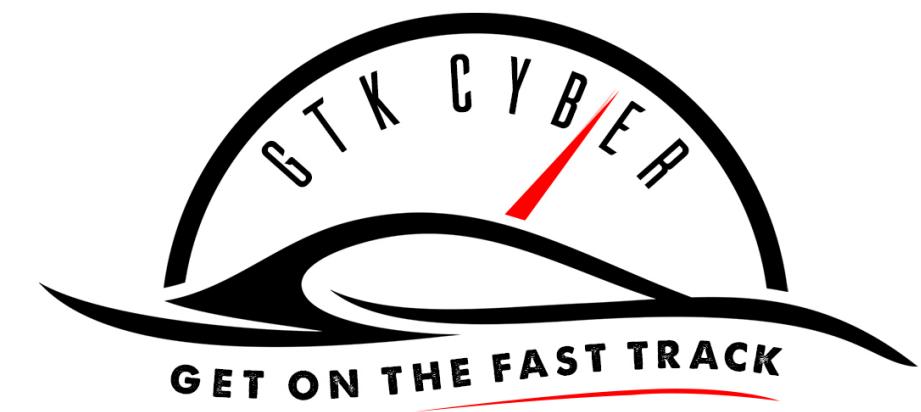
View

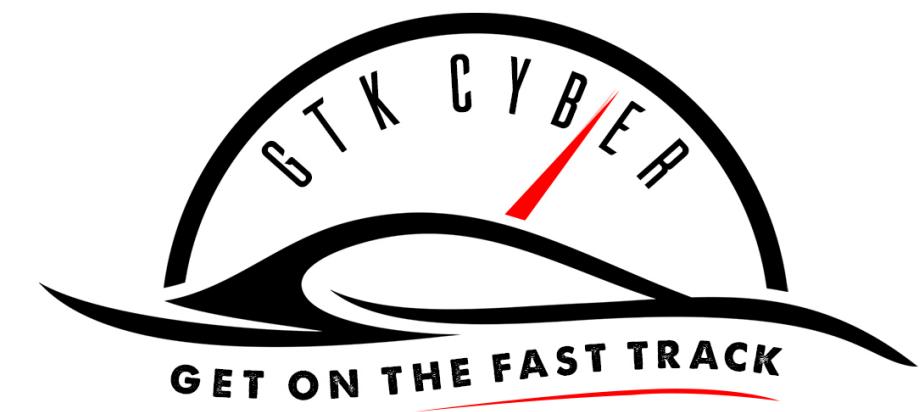
Export

[JSON]

system.indexes

Del



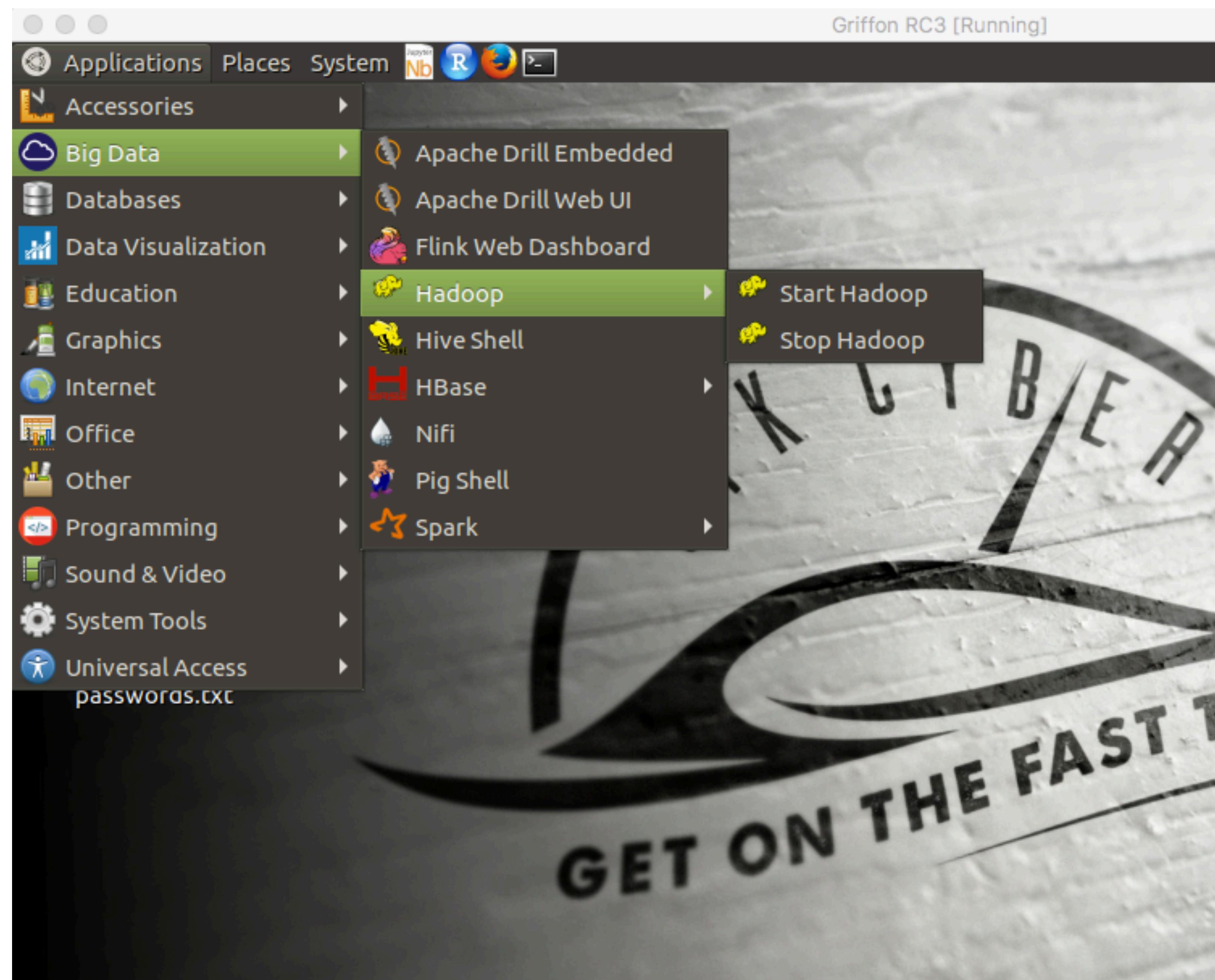


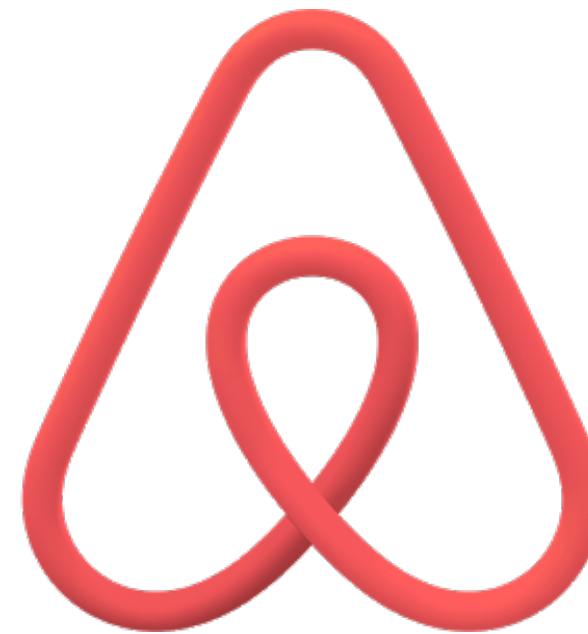
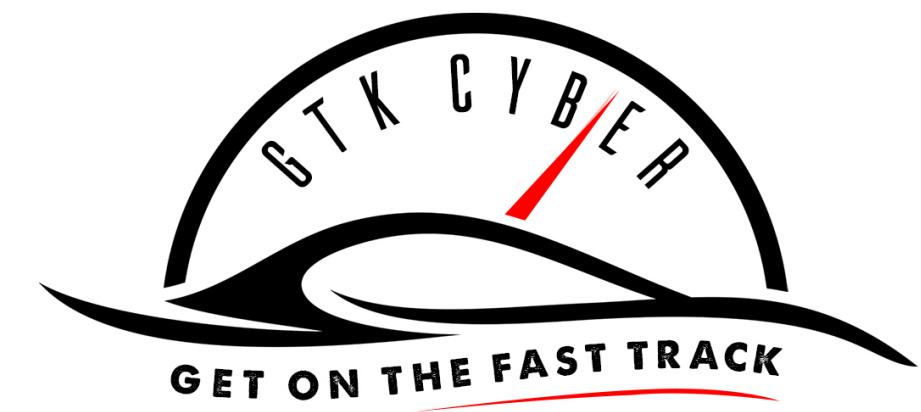
Griffon RC3 [Running]

```
merlin@projectmerlin:/usr/local/hadoop/sbin$ ./start-dfs.sh
16/05/11 23:23:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
merlinuser@localhost's password: 
merlinuser@localhost's password: localhost: Permission denied, please try again.

merlinuser@localhost's password: localhost: Permission denied, please try again.

localhost: Permission denied (publickey, password).
merlinuser@localhost's password:
merlinuser@localhost's password: localhost: Permission denied, please try again.
```

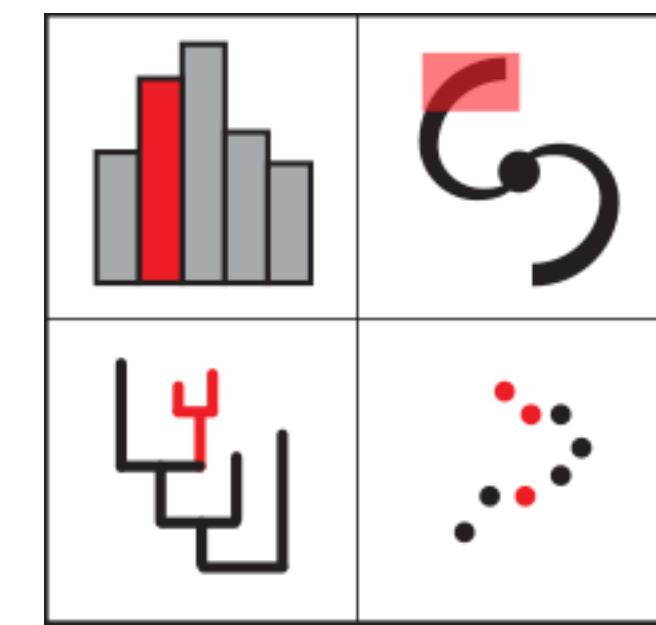


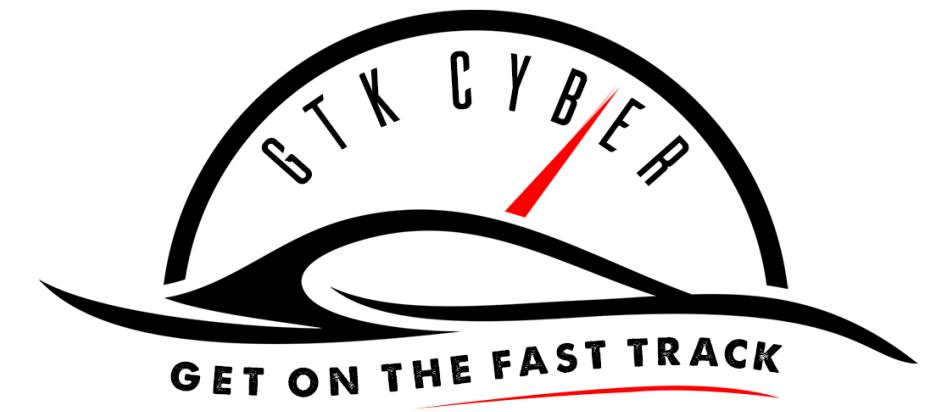


gtkcyber.com



orange
DATA MINING
FRUITFUL&FUN

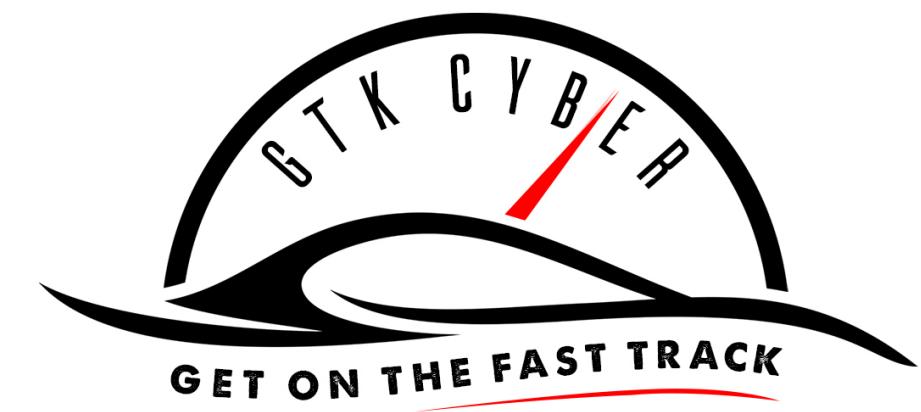




orange

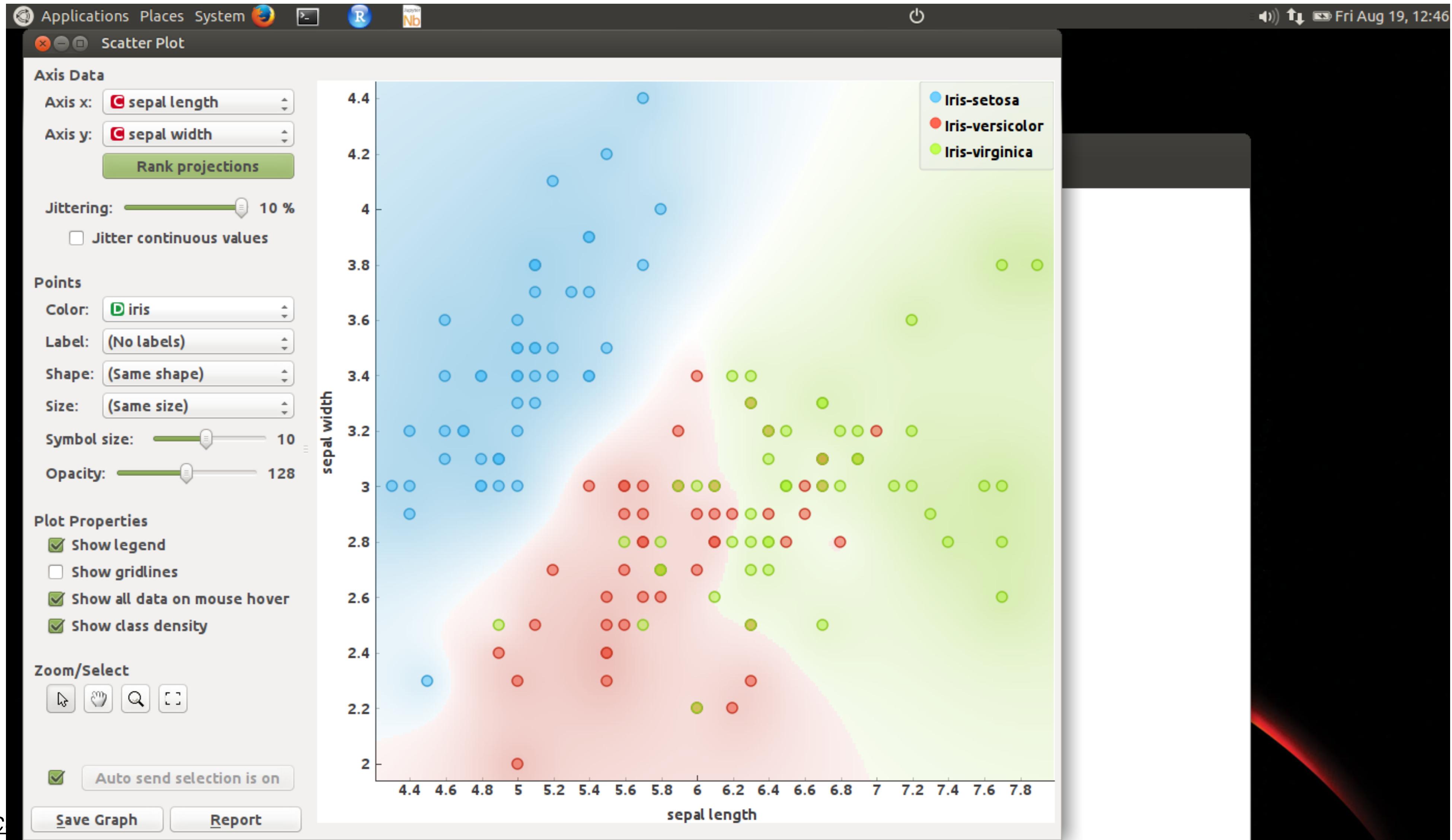
DATA MINING
FRUITFUL&FUN

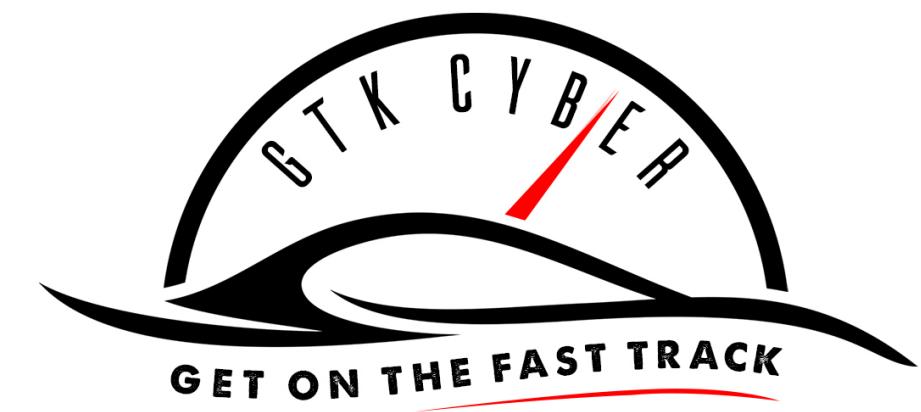




orange

DATA MINING
FRUITFUL&FUN





orange

DATA MINING
FRUITFUL&FUN

Applications Places System R Fri Aug 19, 12:47

Classification Tree Viewer

Tree
13 nodes, 7 leaves

Display

- Zoom
- Width
- Depth
- Edge width
- Target class

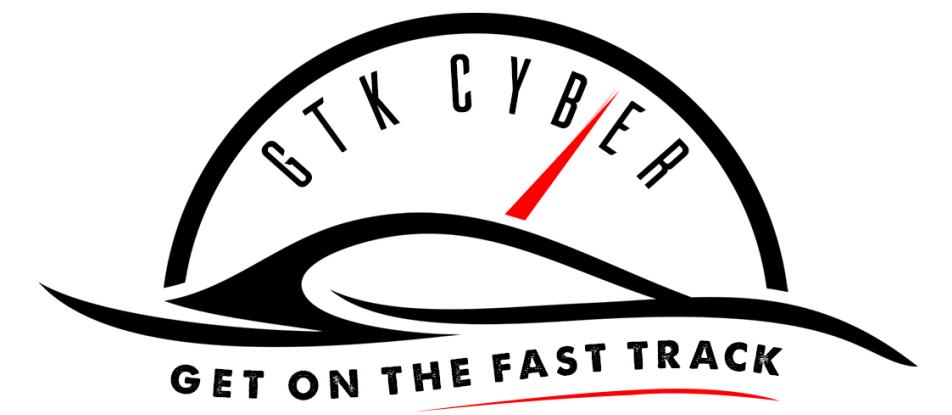
[Save Graph](#) [Report](#)

VirtualBox
Dropped Files

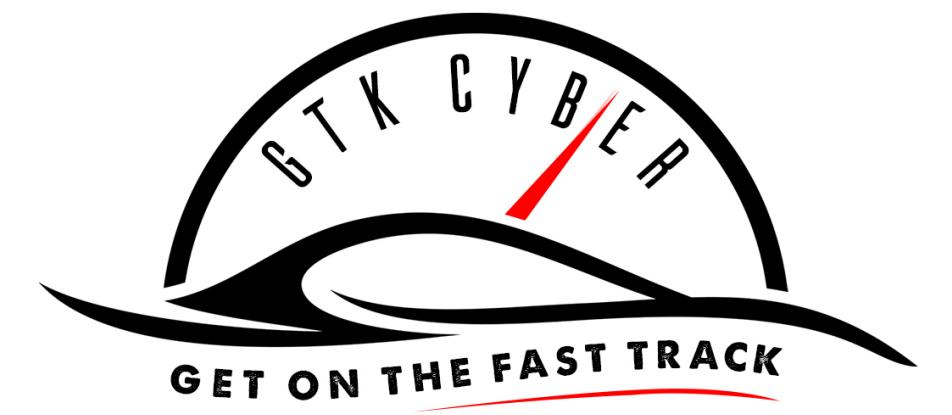
Unsupervised
Prototypes

This workflow works best if you have Tree
Viewer, Scatter Plot and Box Plot all open
at the same time.

the tree viewer
propagates to all the
downstream widgets



Questions?



Using Jupyter Notebook



Using Jupyter Notebook

jupyter

Open a notebook

Files Running Clusters Formgrader Assignments BeakerX Nbextensions

Select items to perform actions on them.

0 / Name

- anaconda3
- Desktop
- Documents
- Downloads
- drill
- metastore_db
- Music
- Pictures
- Public
- snap
- sqlpad
- Templates
- Videos

Upload New

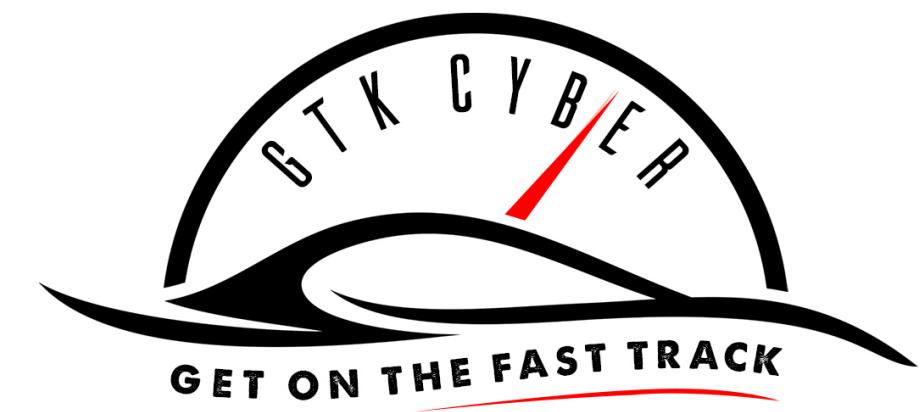
Notebook:

- Bash
- Clojure
- Groovy
- Java
- Javascript (Node.js)
- Kotlin
- PHP
- Python 3
- R
- Ruby 2.5.1
- SQL
- Scala

Other:

- Text File
- Folder
- Terminal

14 days ago



Using Jupyter Notebook

jupyter Untitled Last Checkpoint: 13 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Navigate Widgets Help Snippets Trust

Run

Demo Notebook

This is a markdown cell. It contains the instructions as to how to do the exercises.

In [1]:

```
1 #This is an executable cell
2 print( "This is a python cell")
```

This is a python cell

In []:

```
1 |
```

Run a cell



Using Jupyter Notebook

jupyter Untitled Last Checkpoint: 18 minutes ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Navigate Widgets Help Snippets Trusted Python 3

In [1]:

```
1 #This is an executable cell
2 print( "This is a python cell")
This is a python cell
```

In [2]:

```
1 x = [4,5,6]
```

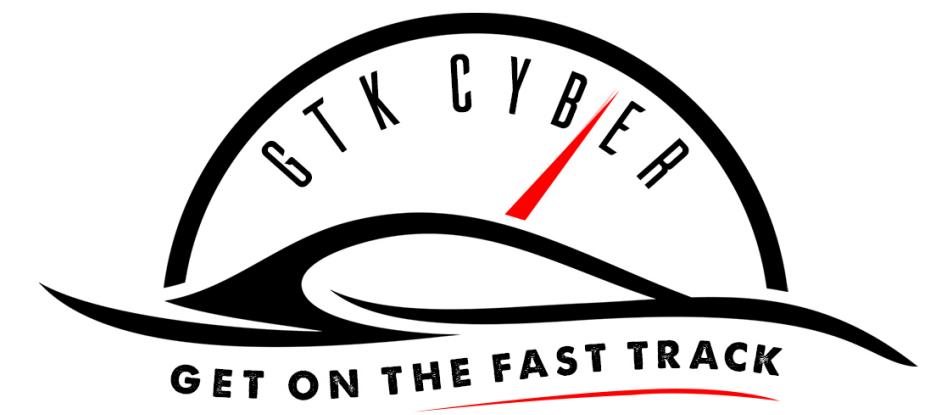
Demo Notebook

This is a markdown cell. It contains the instructions as to how to do the exercises.

Variable Explorer

Variable Inspector

x x list 88 [4, 5, 6]



Questions?

labs.gtkcyber.com

Username: student_0x

Join the Slack Channel at: <http://bit.ly/30lvONs>