

# Python Data Science for Developer



# All About Me Boyd



**Sorratat Sirirattanakarin,  
Boyd**

Data Scientist @3dsinteractive

“Data is a new UX”

## Education

- Master degree, DPU, Big Data Engineering
- Bachelor of Science, B.Sc. (Hons),  
Agro-industry, Biotechnology, KU



DPU

**CITE**  
College of Innovative  
Technology and Engineering  
Dhurakij Pundit University



## ประวัติการทำงานภาคเอกชน

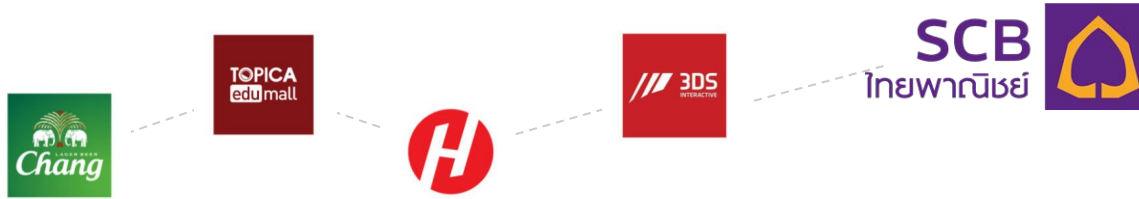
2018 - ปัจจุบันทำงานในตำแหน่ง Data Scientist ในเรื่องของ AI Marketing Automation ณ ที่ บริษัท 3dsinteractive

2017 - ทำงานอยู่ในบริษัท Startup ของต่างประเทศในเครือของ Capstream Ventures Inc ในตำแหน่งของ Digital marketing Executive and Data Scientist (นักการตลาดออนไลน์ นักวิทยาศาสตร์ และวิเคราะห์ด้านข้อมูล)

2016 - Vice Digital Marketing Manager บริษัท Startup จากต่างประเทศ ดูแลเกี่ยวกับการตลาดด้านการศึกษา ผ่าน Social media และ Online Channel

2013 - ประสบการณ์ทำงานในเครือ Thaibev 3 ปี (หัวหน้าฝ่ายปฏิบัติการการผลิตเบียร์ และให้คำแนะนำแก่ฝ่ายการตลาด รวมถึงชุมชน)

2012 - ประสบการณ์ทำงาน Startup กับเครื่องต่างชาติและช่วยในการวิจัย และออกแบบผลิตภัณฑ์ใหม่ (ผู้ช่วยแผนกด้านออกแบบผลิตภัณฑ์ และฝ่ายขาย)



## ประวัติการทำงานภาคสังคม

- อาจารย์ และวิทยากรด้านการตลาดออนไลน์ ที่บริษัท "siamhr" จัดสอนและอบรมให้กับทางระดับผู้จัดการ ผู้อำนวยการ และผู้ประกอบการ

- อาจารย์ และวิทยากรด้านการตลาดออนไลน์ ให้กับผู้ประกอบการที่ สำนักงานพาณิชย์ จังหวัดปทุมธานี และชมรม BizClub จังหวัดปทุมธานี

- กรรมการด้าน สารสนเทศและเทคโนโลยี (การตลาดออนไลน์) และให้คำแนะนำเพื่อสร้างความแข็งแกร่งให้กับผู้ประกอบการไทย ที่ "สมาคมนักธุรกิจอาเซียน"

- สมาชิกกลุ่ม BKK AI Volunteer มอบความรู้ด้าน AI ให้กับผู้ที่เริ่มต้นไปจนถึงขยายความรู้ด้าน AI ให้กว้างขวางยิ่งขึ้น



**FANPAGE - [www.facebook.com/bigdatarpg](http://www.facebook.com/bigdatarpg)**



BigData RPG  
@bigdatarpg

Home

Live Videos

Videos

Reviews

Posts

Photos

About

Community

Events

Notes

Info and Ads



[Edit Page Info](#)
[View as Page Visitor](#)
[View Insights](#)
[...](#)

[Create Post](#)
[Live](#)
[Event](#)
[Offer](#)
[Job](#)

 Write a post...

[Photo/Video](#)
[Feeling/Activ...](#)
[Write Note](#)
[...](#)

[Share a recent photo from your Instagram account](#)

[Only you can see this](#)



[Learn More](#)

Personal Blog in Bangkok, Thailand  
 5.0 ★★★★★  
 Always Open



ABOUT BIGDATA RPG



# Afternoon

---

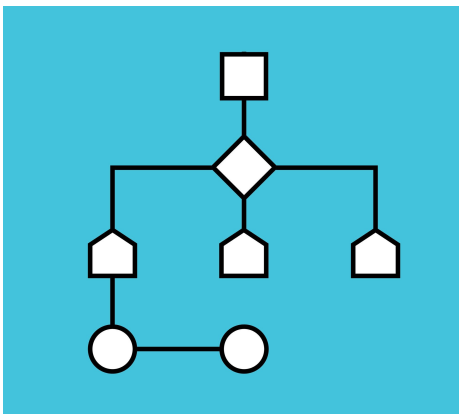
Day-01



Boyd BigData RPG - Sorratat Sirirattanajakarin

1. Kaggle Extension
  - Linear Regression Modeling

# เข้าใจศัพท์ก่อนลุย



Algorithm

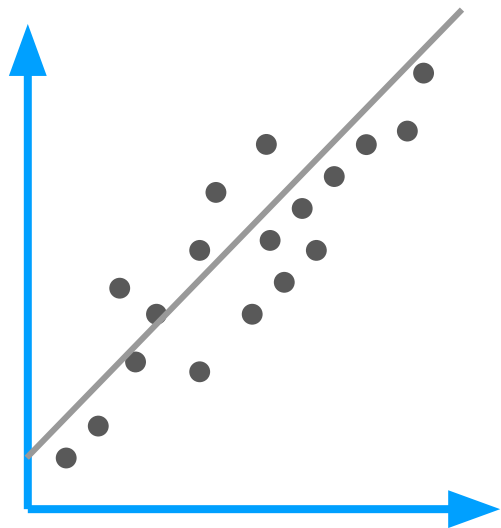


Data



Model

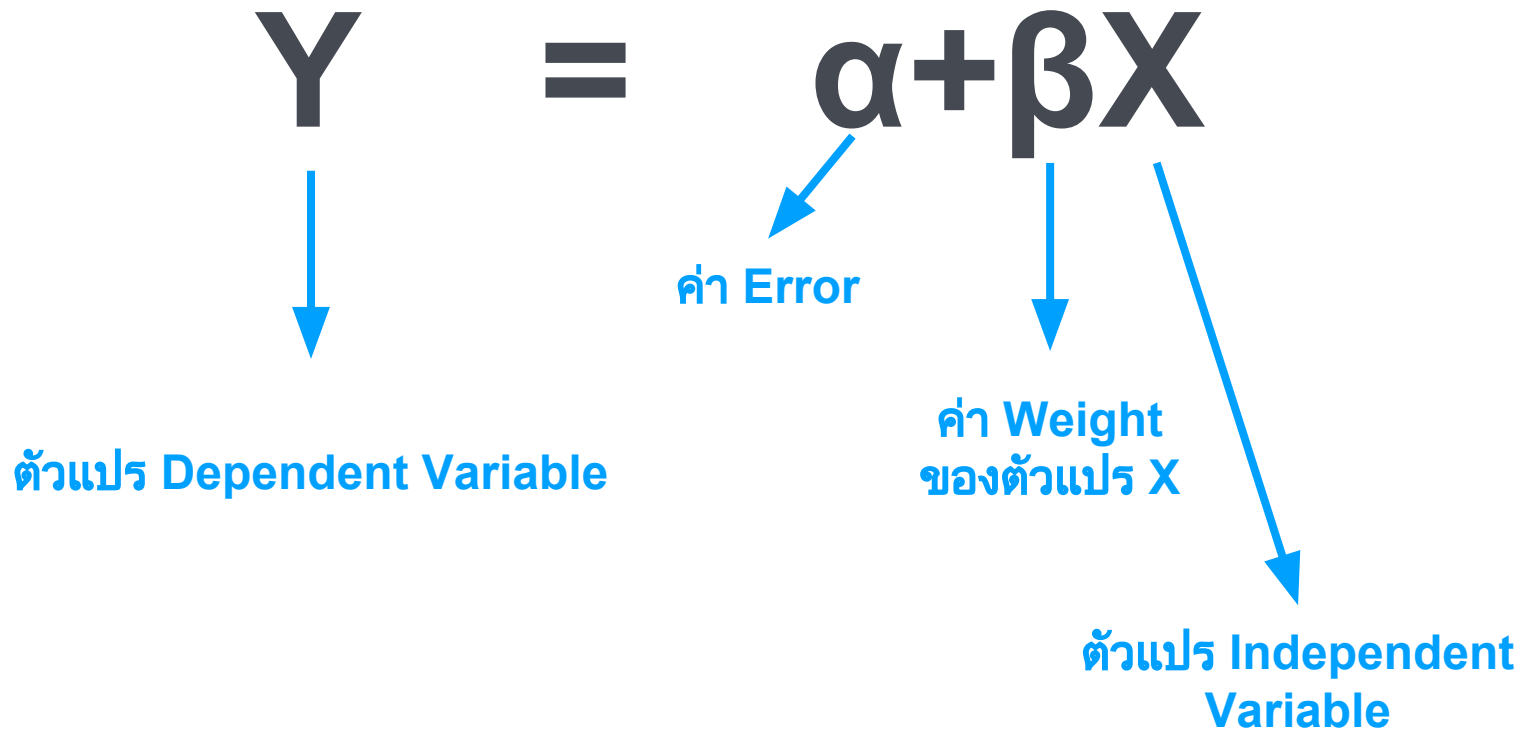
# Linear Regression



Linear  
Regression

สมการเส้นตรงสุดแสดจะธรรมดา ก็เกิดขึ้นมาเพื่อการทำ Machine learning ในสายของข้อมูลตัวเลข เพราะทุกๆจุดสามารถเป็นคำตอบได้ทั้งหมด และข้อมูลที่ส่งเข้าไปเรียนก็เป็นกลุ่มของตัวเลขด้วยเช่นกัน

# Linear Regression





# Map formula to Data set

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

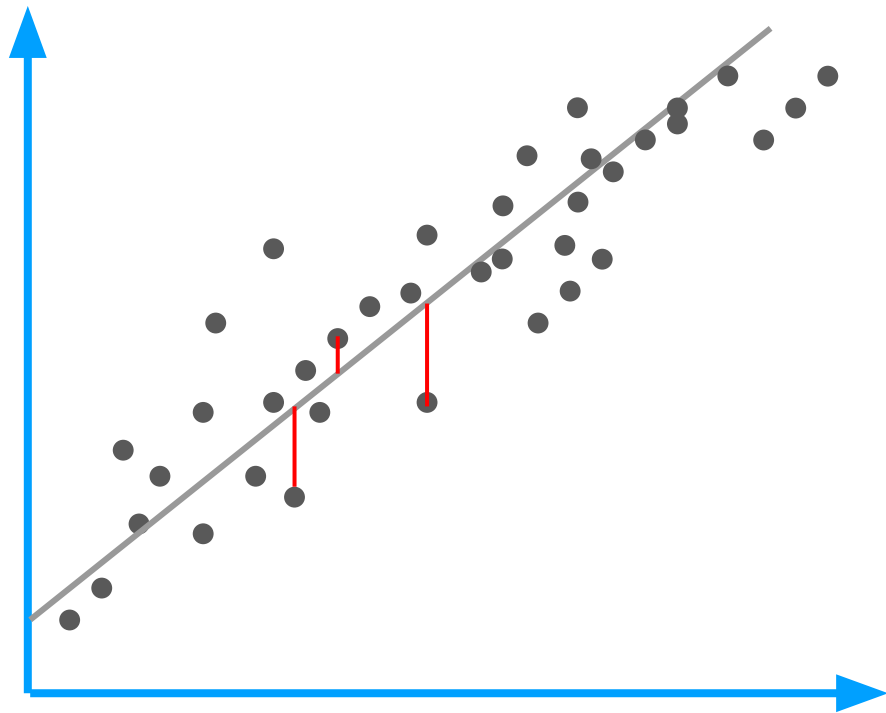


Label



Features

# Error Evaluation



ผลรวมความห่างของจุดต่างๆ กับเส้นที่  
พาดผ่านทั้งหมดยกกำลังสอง เราเรียกว่า

ค่า Sum Square Error

$$SSE = \sum (y - y')^2$$

# Error Evaluation

RMSE คือค่าที่ไว้วัดความแม่นยำของ Model โดยดูจาก Error ที่เกิดขึ้น ดังนั้นหากค่านี้น้อยแสดงว่ามี Error น้อย นั่นหมายถึงความแม่นยำของ Model มากนั่นเอง หรือแต่ละจุดบนกราฟ ล้วนเข้าใกล้เส้นที่พาดผ่าน

$$\text{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

# Install Package

```
conda install scikit-learn  
pip install scikit-learn
```

1

2

หากขั้นแรกไม่ได้ให้พิมพ์ขั้น 2  
แทนในการติดตั้ง

# Kaggle Competition

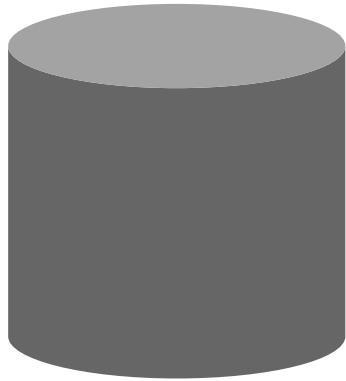


<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/leaderboard>



# Kaggle Competition

---



**Training  
Data**



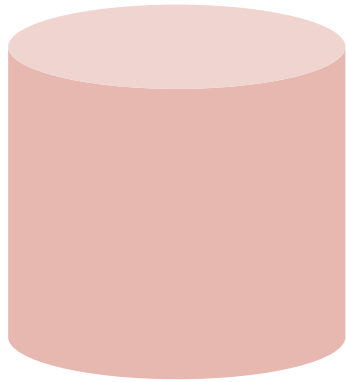
**Testing Data**



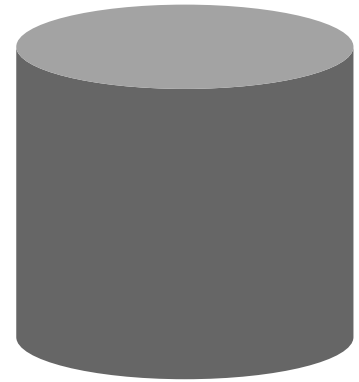
**Submission  
Data**

# Kaggle Competition

---



**Raw Data**



**Train Data**

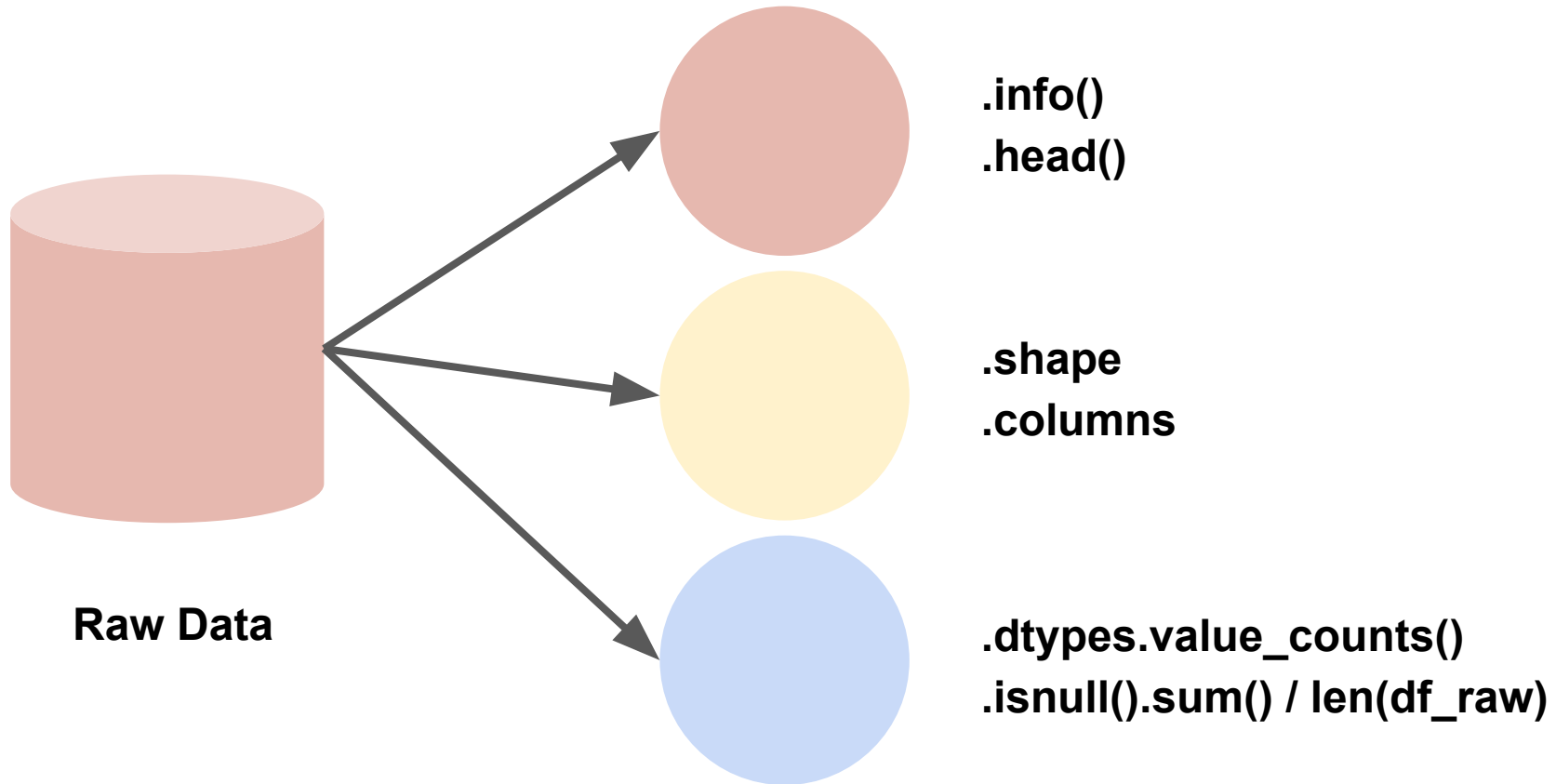
# Kaggle Competition

---

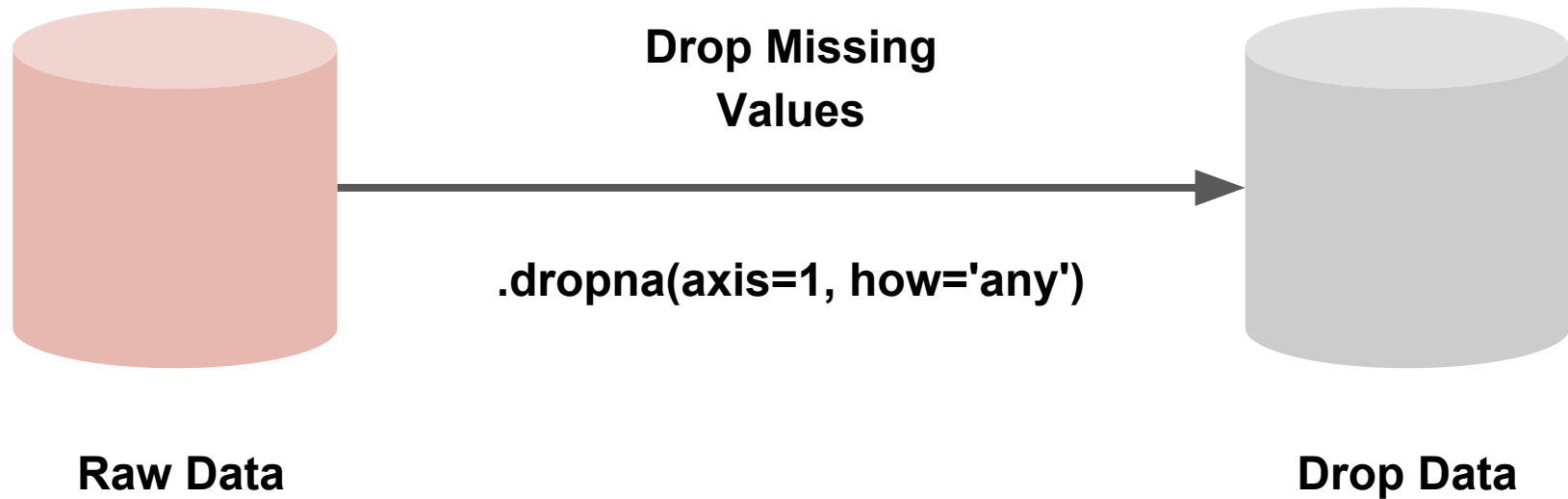




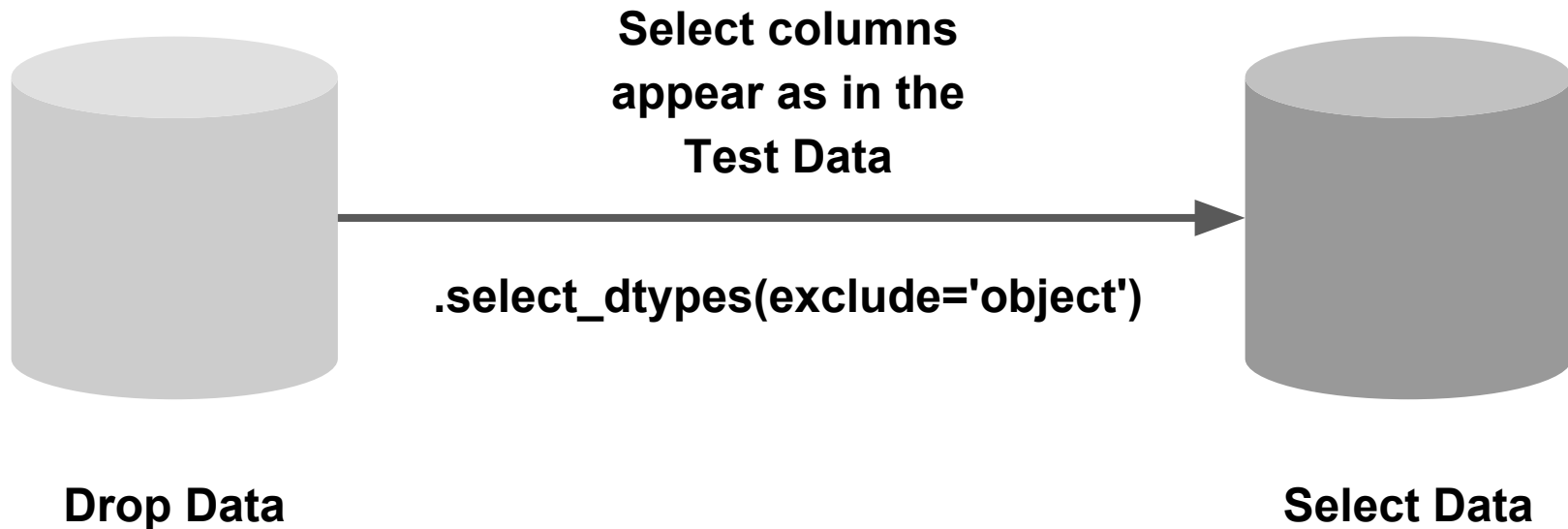
# Kaggle Competition



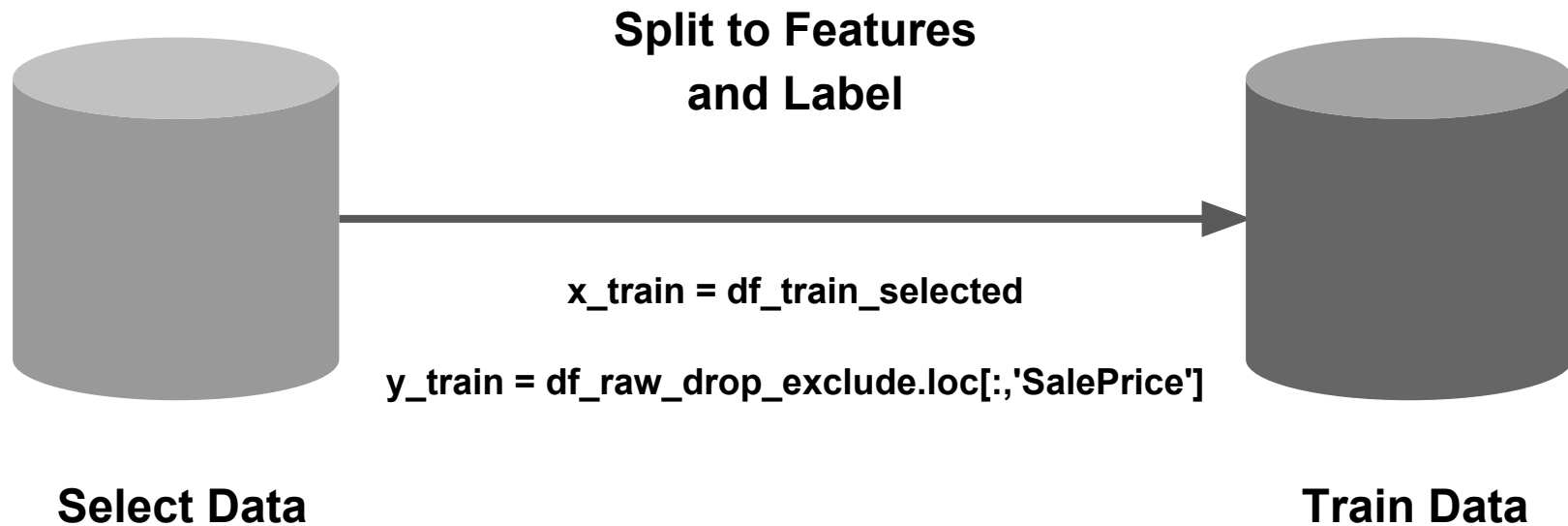
# Kaggle Competition



# Kaggle Competition



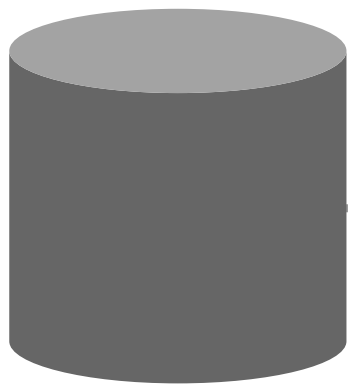
# Kaggle Competition



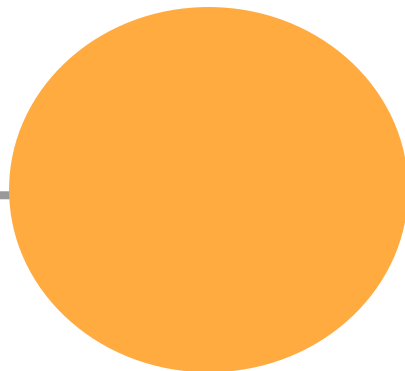
# Kaggle Competition



3 ส่วนใหญ่ที่เราทำ Modeling



**Train Data**



**Modeling**



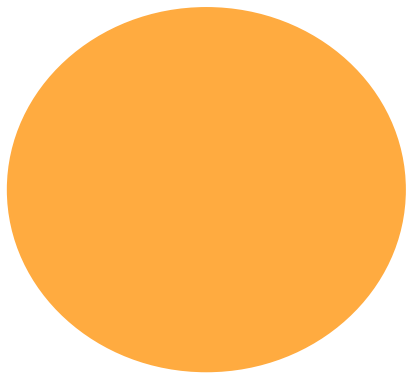
**Evaluation**

# Kaggle Competition

---



## Modeling

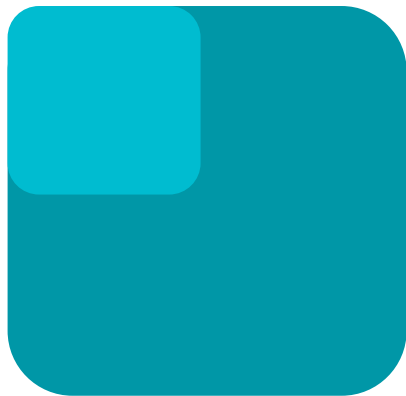


- ทำการเรียก Packages  
`from sklearn import linear_model`
- ทำการตั้งค่าให้กับ Model linear\_model
- ทำการ Train โดยใช้ Method .fit(X,Y)
- เรียบร้อยเราได้ Model ที่พร้อมทดสอบแล้ว !!

# Kaggle Competition



## Evaluation



- ในที่นี้เนื่องจากคำตอบเราเป็น Continuous Value
- เราจึงใช้ RMSE ในการวัดความถูกต้องของข้อมูลว่าข้อมูลที่ทำนายคลาดเคลื่อนจากข้อมูลจริงๆแค่ไหน
- RMSE (Root Mean Square Error)

อย่าลืมส่งคำตอบกันนะ !!

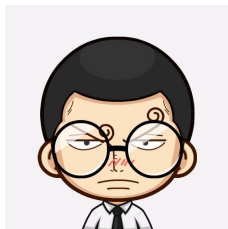
= ]



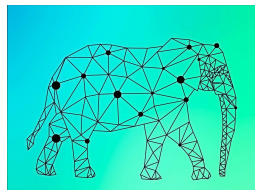
# Keep more practice



DataCube



DataRockie



“Let the data tell story”

ชี้ช้างจับข้อมูล



LinkedIn



DataCamp



DataCube



Kaggle

# Python Data Science for Developer

Thank you

