Spak

Organization

Optimization

Statistics

Data exploration

Prediction

Data structure( unsupervised)

Optimization( operations reserarch)

e.g

exploration:

Chicago！ How many zip?

Factors? Scope: de

Visualization plot: geo 2d3? Tableow?

weather

Economic indicators S&P 500

Cloud experience

Ggplot2 python package

One model predicting another model explaining

Odds: p/1-p = eη(x) 逻辑回归 logistic regression test the model good or not

Profit by facebook???

BCNF unjoint, minimize database?

3NF??? Join database

2nd meeting

Goal:

How to store data: data structure: BCNF or 3NF. 使用数据库 数据结构

(K,V) primary key, secondary key

Energy consumption (zip, value)----realationship(zip,date)-----> factors

1--- many or many---many DB schema sigma—calculator(SQL)

Cencus---zip relationship

For energy consumption : energy use by zip or by census

ISA?? Relation model ISA

Agg data??

GCP: google cloud platform: severless: data flow bigQuery

Spak (databvicks, conmunity env)

Signal platform?

Weather data: NLAA geological

More about commercial business type

Plan: EDA(exploring data analysis), plot,ggplot? joining data

CRISP-DM framework??

GCP sign up data engineering certification?

Data architech certification? IAM xx access management

AWS Azuc??

CAP?? Consistence available

Third notes:

WN/ non-WN

Block/randomize

Piecewise linear model

Sql lite

Google cloud database

Big query

Reighley correlation

Linear model

Correction term building