

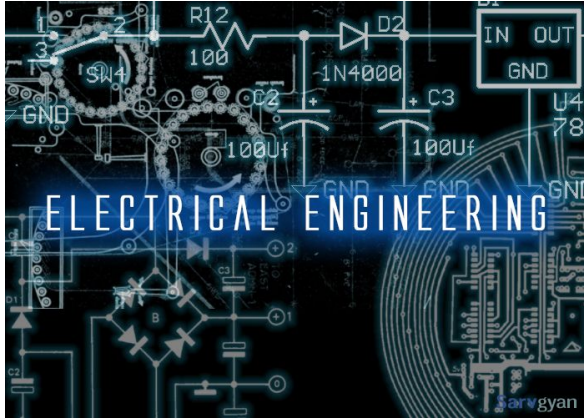
# Hello!

## I am Zhian Wang

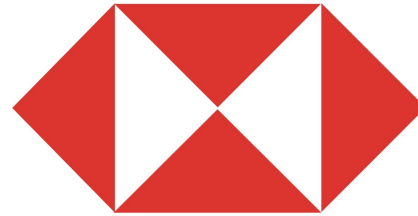
---

I am here because I love data and analytics.

# About me



BE '13,  
Electrical Engineering,  
SMU, China



# HSBC

MS '18,  
Business Analytics,  
GWU, US



# GWU MSBA

- R, Python, SQL, SAS, AMPL
- Data Mining, Data Management
- Statistics, Time Series, Probability Models
- Optimization, Decision and Risk Analytics
- ...
- GPA: 3.92/4.0 (12 A and 4 A-)
- Teaching Assistant for 2 semesters

# Projects and Internships

AUG 16

- Oscar Best Picture Nomination
- Austin Service Request Analysis
- 2016 US Presidential Election Swing States Analysis
- Customer Retention Analysis
- Top 100 Restaurants in DC

- World Bank - Data Viz Gallery
- Kaggle Competitions
- Norway New Car Sales Forecasting

JAN 17

AUG 17

WRI  
Project  
Database  
Intern

CSIS

Analytics Intern

SEP 17

# Analytics

- 2017 Oscar Best Picture Nomination
- Kaggle Competitions
- Norway New Car Sales Forecasting

# Visualization

- Interactive Time Series Line Chart
- CSIS - Table App
- CSIS - Chart Maker

# Analytics

# 2017 Oscar Best Picture Nomination

- Data Collection

Web Scraping - IMDB, Rotten Tomatoes (2012-2016)

Sentiment Analysis - Twitter

A Kaggle Movie dataset

- Data Consolation - PostgreSQL

- Logistics Regression - R

- Demonstration - Shiny

Data Exploration

Regression Model Result

# Prediction of 2017 Oscar Nomination

**File Input**

Browse...

movie.csv

Upload complete

**Training Data**

2,012

2,014

2,015

2,012

2,013

2,014

2,015

**Validation Data**

2,012

2,015

2,012

2,013

2,014

2,015

## Summary

[DataSummary](#)[Stepwise Model Summary](#)[Logistic Model Summary](#)[Confusion Matrix](#)[Misclassification Rate](#)[ROC Curve](#)[Lift Curve](#)

Nomination Movies for 2017 Oscar

### data

Finding Dory

Zootopia

Doctor Strange

Manchester by the Sea

Moonlight

Aquarius

The Wailing

La La Land

*Animated Feature Film*



# Sberbank Russian Housing Market - Top 12%

Predict realty prices based on housing data and macroeconomic patterns

## Weighted Averaging

- 3 XGBoost models
- Data Processing
  - Remove extreme values
  - Handle unreasonable value
  - Add new features
  - Encode categorical variables
  - Deflate the house price

## Stacking (H2O)

- Data Processing
  - Remove extreme values
  - Handle unreasonable value
  - Add new features
  - Encode categorical variables
  - Combine features
- Random Forest
- Extreme Random Tree
- GBM
- XGBoost

# Norway New Car Sales Forecasting

- Monthly sales data from Jan 2007 to Jan 2017
- Two variables: Quantity and Import

Model		RMSE	Variance
Univariate	IMA(1,1)	1,074.8	1,245,614
	Linear Trend + Seasonal Dummies +Drop + AR(3)	1,180.4	982,806
	Simple Exponential Smoothing	1,072.5	1,234,014
Bivariate	TF b=0 s=1 r=2 with noise model MA(1)(12)	NA	760,830
	TF b=0 s=1 r=1 with noise model MA(1)(12)	NA	814,682
	Intervention (b=0,s=2,r=1, MA(1)(6) error model)	NA	1,073,179

Model	Model Variance
TF b=0 s=1 r=2 with noise model MA(1)(12)	760,830
Linear Trend + Seasonal Dummies +Drop + AR(3)	1,012,732

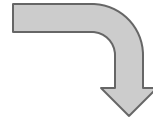
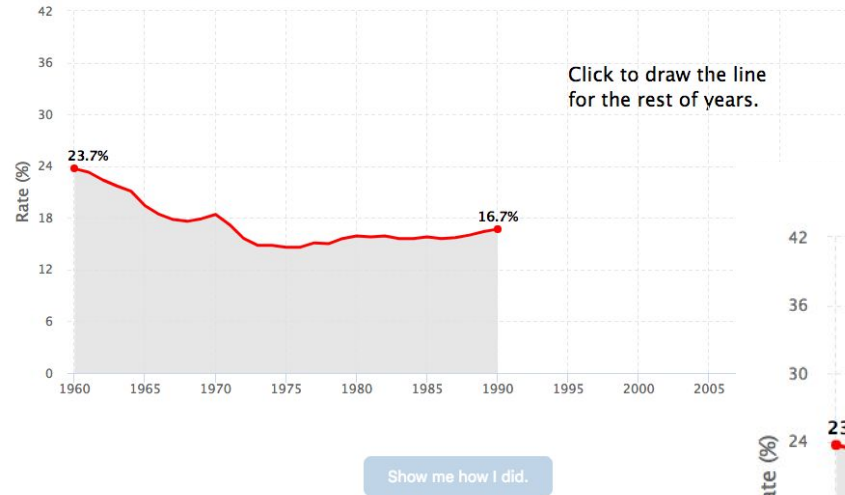
# Visualization



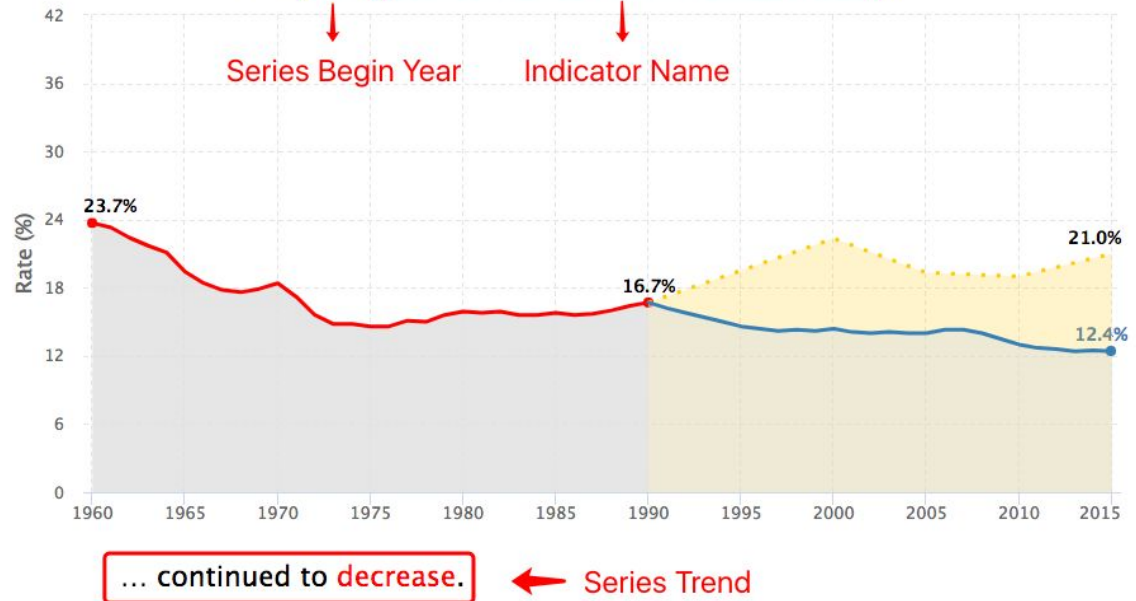
# World Bank - Data Visualization Gallery

## WDI Time Series Interactive Line Chart

Since 1960, Birth rate, crude (per 1,000 people) ...



Since 1960 Birth rate, crude (per 1,000 people) ...



Apply the line chart to thousands of WDIs

# CSIS - Table App

CSIS

CENTER FOR STRATEGIC &  
INTERNATIONAL STUDIES

INTERNATIONAL  
SECURITY PROGRAM

Choose a dataset

FPDS\_ContractType

View

☒ Standard ☐ Difference

Aggregate by

- ☒ Customer  
☒ Category  
☐ Contract.Type  
☒ VendorSize  
☒ Portfolio  
☐ Classification

Copy

Print

Download

Search:

Fiscal Year

Customer

Category

VendorSize

Portfolio

Amount

All

["Navy", "Air Force"]

A

["Big Five"]

All

All

1	2000	Air Force	Products	Big Five	Aircraft and Drones	\$11,737,531,132
2	2001	Air Force	Products	Big Five	Aircraft and Drones	\$13,390,540,520
3	2002	Air Force	Products	Big Five	Aircraft and Drones	\$15,362,432,386
4	2003	Air Force	Products	Big Five	Aircraft and Drones	\$16,197,836,544
5	2004	Air Force	Products	Big Five	Aircraft and Drones	\$14,648,406,856
6	2005	Air Force	Products	Big Five	Aircraft and Drones	\$11,357,578,661
7	2006	Air Force	Products	Big Five	Aircraft and Drones	\$16,305,378,294
8	2007	Air Force	Products	Big Five	Aircraft and Drones	\$17,634,881,372
9	2008	Air Force	Products	Big Five	Aircraft and Drones	\$12,972,229,204

# CSIS - Table App

Choose a dataset

User Select

Please choose a csv file to read, which should have fiscal year in the 1st column and an 'Amount' column.

Browse...

FPDS\_Customer\_data.csv

Upload complete

View

☐ Standard

☒ Difference

Aggregate by

☒ VendorSize

☒ Customer

☒ Category

☒ Portfolio

☐ Contract.Type

☐ Classification

Difference between

FY\_2015

and

FY\_2016

Filter Zero Data

☐ Yes

☒ No

Show 50 entries

Search:

	VendorSize	Customer	Category	Portfolio	FY_2015	FY_2016	Difference	Percent_Difference
	All	All	All	All	All	All	All	All
1	Big Five	Air Force	Products	Aircraft and Drones	\$7,845,625,956	\$15,229,575,770	-\$7,383,949,814	48%
2	Big Five	Air Force	Products	Electronics and Communications	\$1,115,581,272	\$1,264,744,290	-\$149,163,018	12%
3	Big Five	Air Force	Products	Facilities and Construction	\$861,634	\$4,706,062	-\$3,844,428	82%
4	Big Five	Air Force	Products	Land Vehicles	\$17,326	-\$11,379	\$28,705	-252%
5	Big Five	Air Force	Products	Missile and Space Systems	\$1,130,877,105	\$2,080,996,521	-\$950,119,416	46%
6	Big Five	Air Force	Products	Other Products	\$560,380	\$0	\$560,380	100%
7	Big Five	Air Force	Products	Ships & Submarines	-\$24,204	\$39,000	-\$63,204	-261%
8	Big Five	Air Force	Products	Weapons and Ammunition	\$1,734,239,112	\$2,468,665,529	-\$734,426,417	30%

# CSIS - Chart Maker

## FPDS Charts

Years



- ☒ Bar Chart  
☐ Line Chart  
☐ Double Stacked

Breakout

PlatformPortfolio

Facet

None

Y Variable

Action.Obligation.2016

- ☐ As Total  
☒ As Share

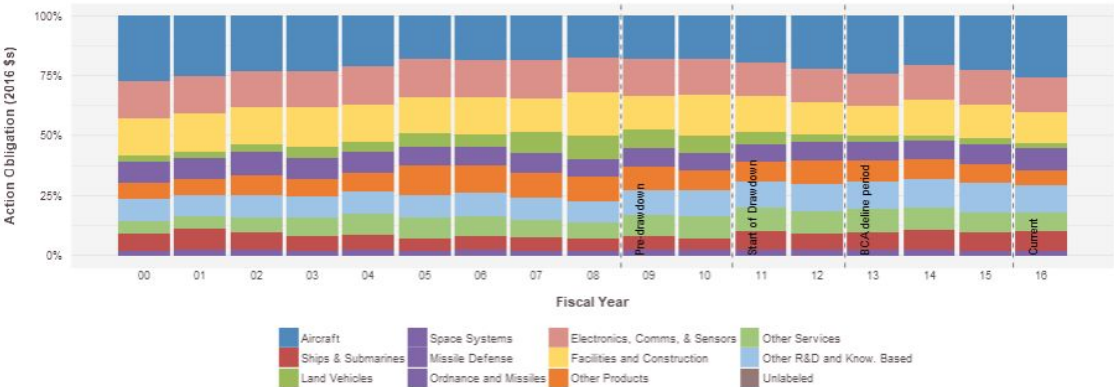
Show Drawdown Periods

- ☒ Yes  
☐ No

Charts

Edit Data

Share of Action.Obligation.2016 by PlatformPortfolio



Save Plot as JPG (300 DPI)

Height (inches):

6

Width (inches):

6

Download Plotted Data

Plot title

Share of Action.Obligation.2016 by PlatformPortfolio

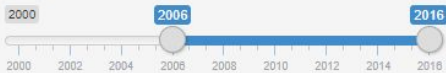
☒ Show title

Lock

# CSIS - Chart Maker

## FPDS Charts

### Years



- ☐ Bar Chart
- ☐ Line Chart
- ☒ Double Stacked

### Breakout

SubCustomer

### Facet

Shiny.VendorSize

### Y Variable

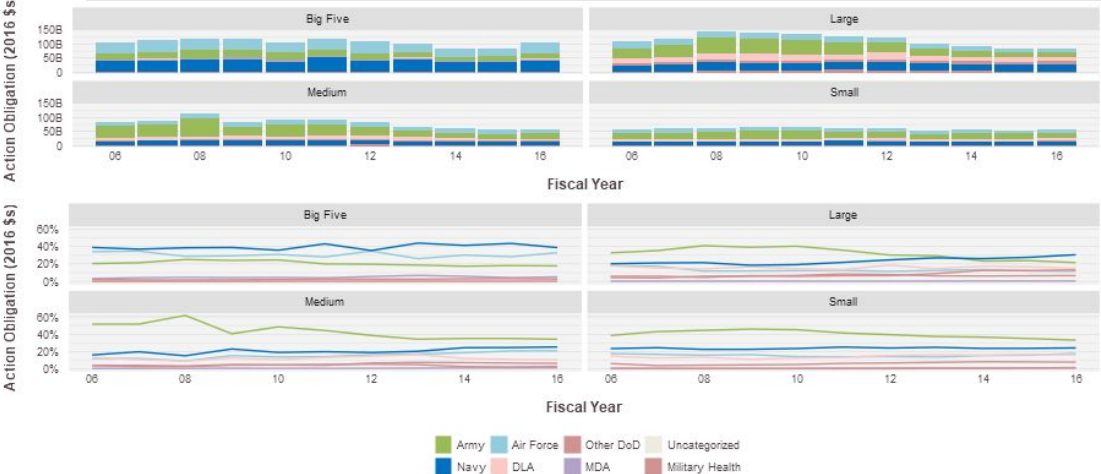
Action.Obligation.2016

### Show Drawdown Periods

- ☐ Yes
- ☒ No

### Charts

### Edit Data



Save Plot as JPG (300 DPI)

Height (inches): 6

Width (inches): 6

Download Plotted Data

### Plot title

Share of Action.Obligation.2016 by SubCustomer and Shiny.VendorSize

Show title

Lock



# Thanks!