OPEN DATA SCIENCE CONFERENCE

#ODSC

@ODSC

Boston | May 1 - 4 2018

zilliant

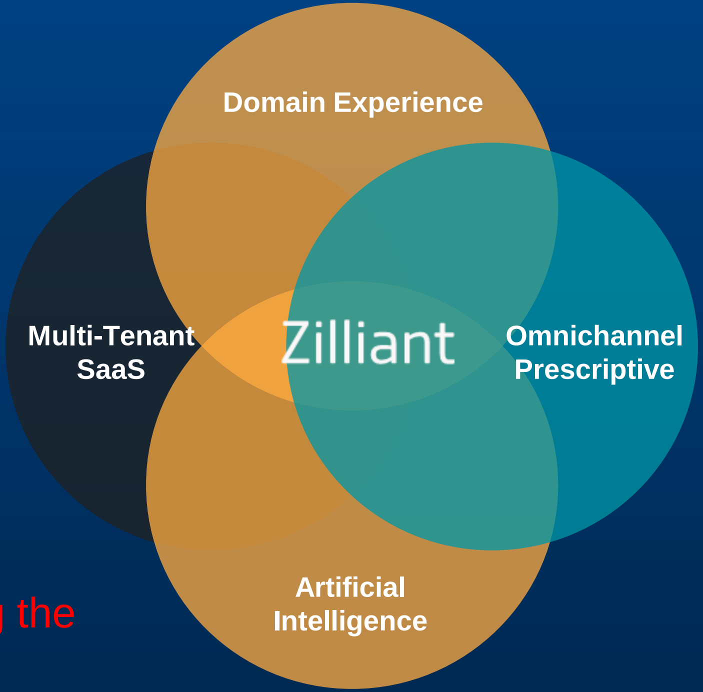
Bayesian Hierarchical Model for Predictive Analytic

Amir Meimand
Director of R&D, Pricing Science, Zilliant

zilliant

About Zilliant

- Founded in 1998
- Headquartered in Austin, Texas
- 200 employees & contractors
- Serving B2B Distribution, Manufacturing & Industrial Services
- 120+ customers around the world. 100’s of implementations
- Investors: ABS Ventures, Houston Ventures, Trellis Partners,
- and Goldman Sachs



The world’s leading AI-enriched SaaS platform for maximizing the lifetime value of B2B customer relationships.

About Zilliant

Zilliant IQ

Converting Strategic Insights into Account Specific Action Plans

Ex. Data In

- Product Master
- Customer Master
- Transaction Data
- Win / Loss Data
- 3rd Party Data
- Competitor Price Data

Zilliant IQ – Multi-Tenant SaaS Platform

Cloud Infrastructure: Highly Scalable – Secure – Collaborative – In-Memory Reporting

Action IQ – Actionable Intelligence & Sales Guidance

- Account Potential
- Account Health
- Dashboards & Collaboration
- Action Plans & Campaigns
- Quotes
- Agreements

IQ Engines – AI & Prescriptive & On-Demand Insights

- Cart IQ
- Cost IQ
- Price IQ
- Profit IQ
- Sales IQ

Zilliant IQ Anywhere – Integration & Omnichannel Activation

Omnichannel Activation

Zilliant Apps.

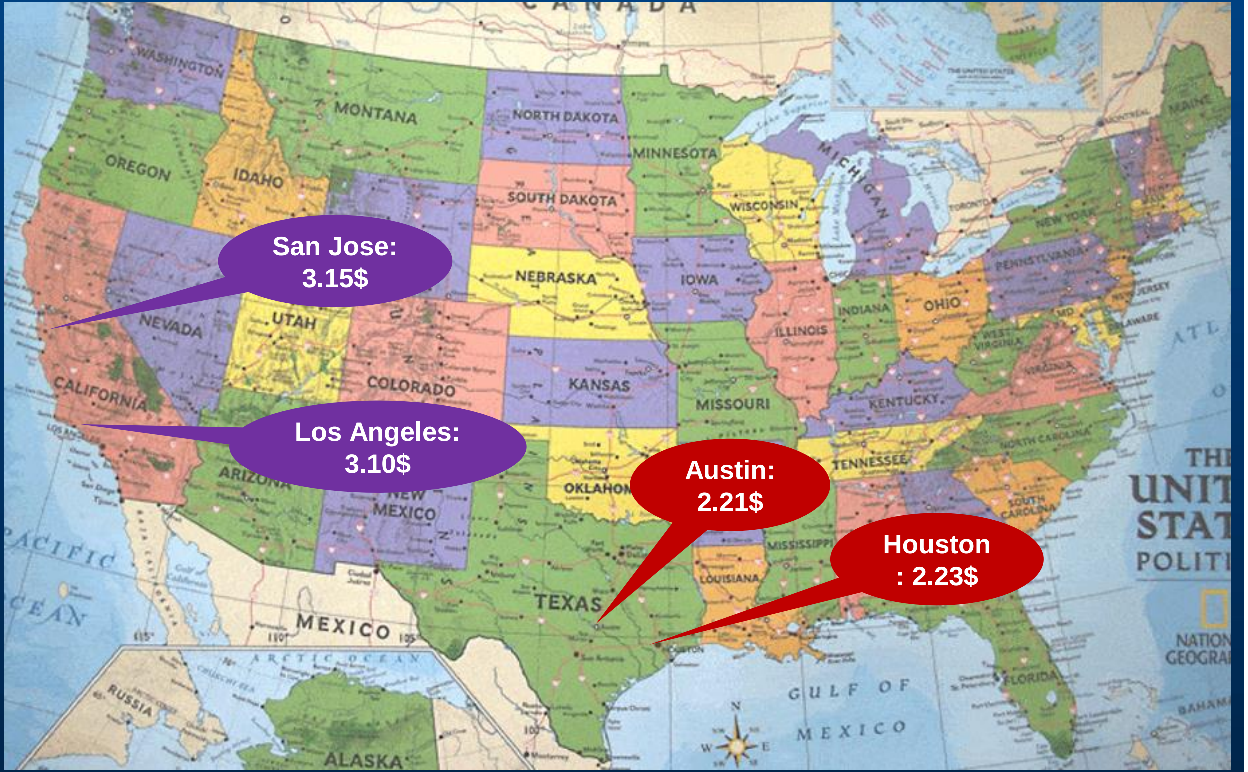
- Email
- Mobile
- Browser

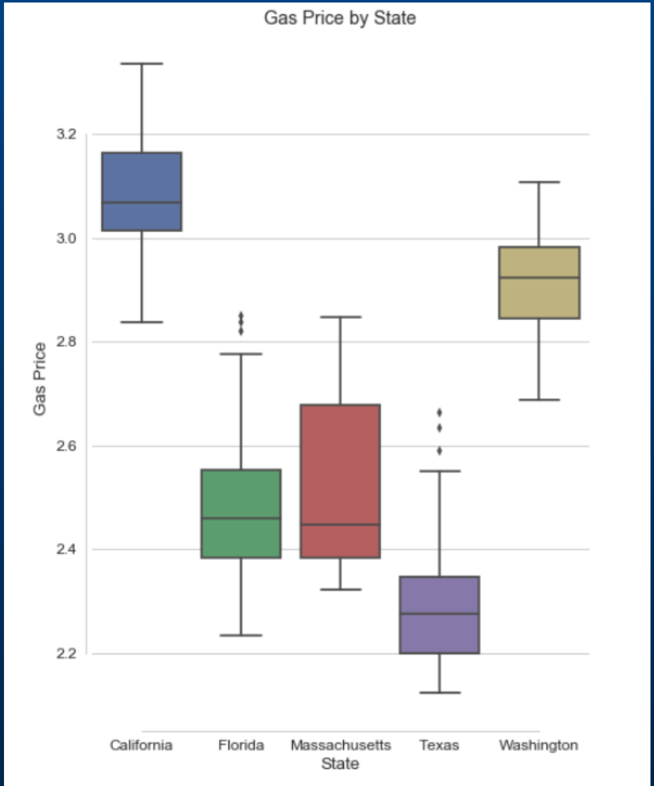
Enterprise Integration…

- CRM / CPQ
- eCom
- Quoting Tools
- ERP

zilliant

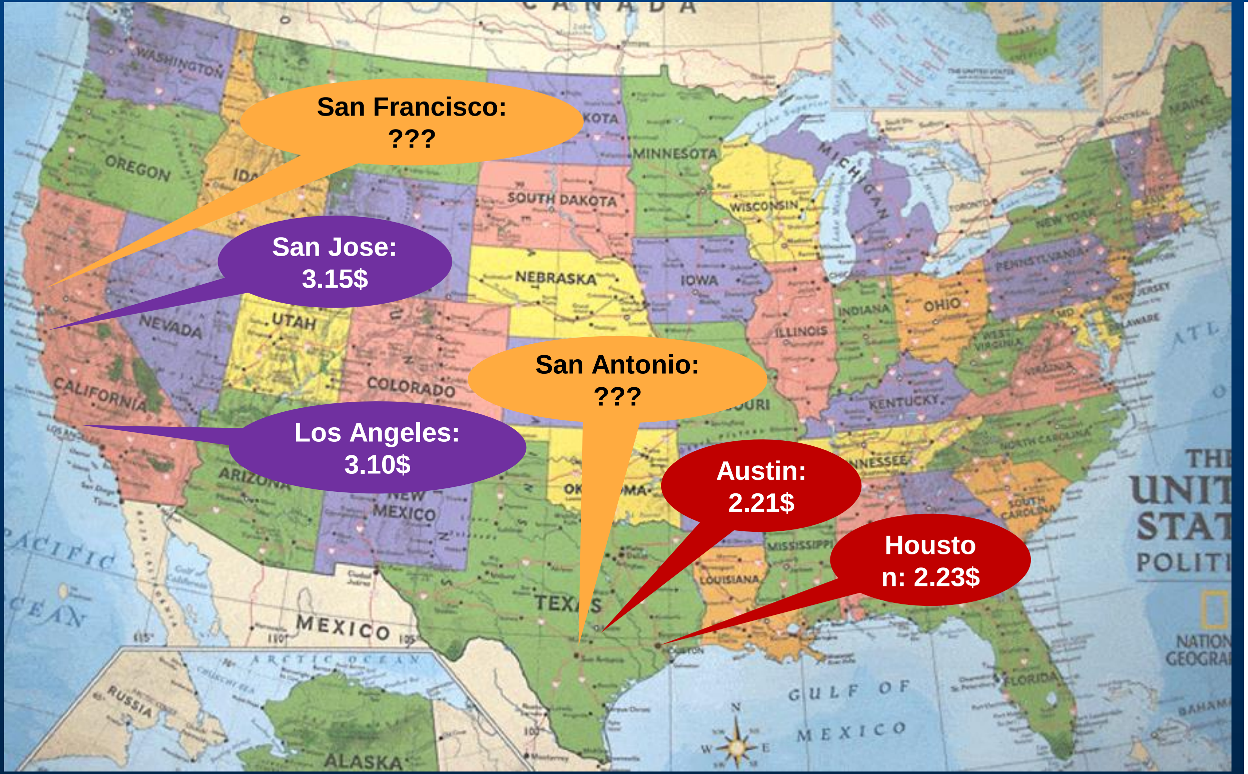
Gas Price:

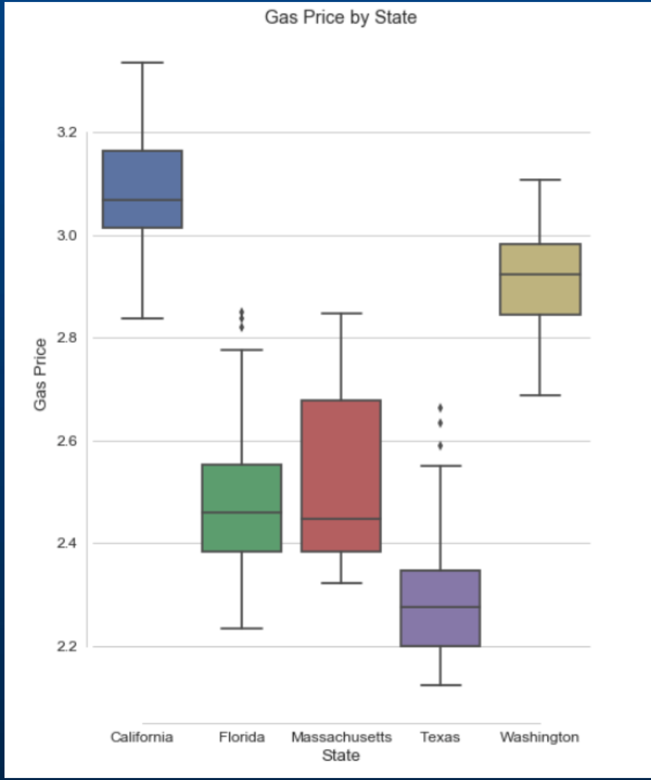




Source: https://www.eia.gov

Gas Price:

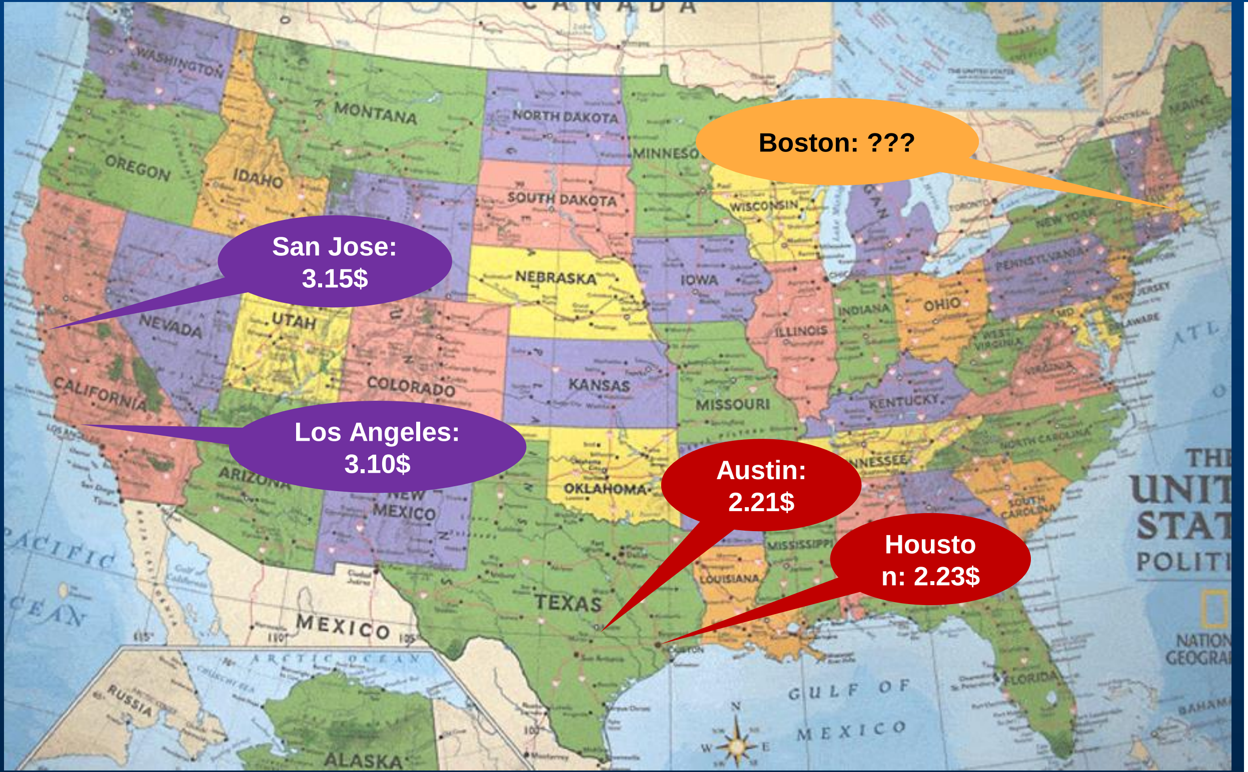


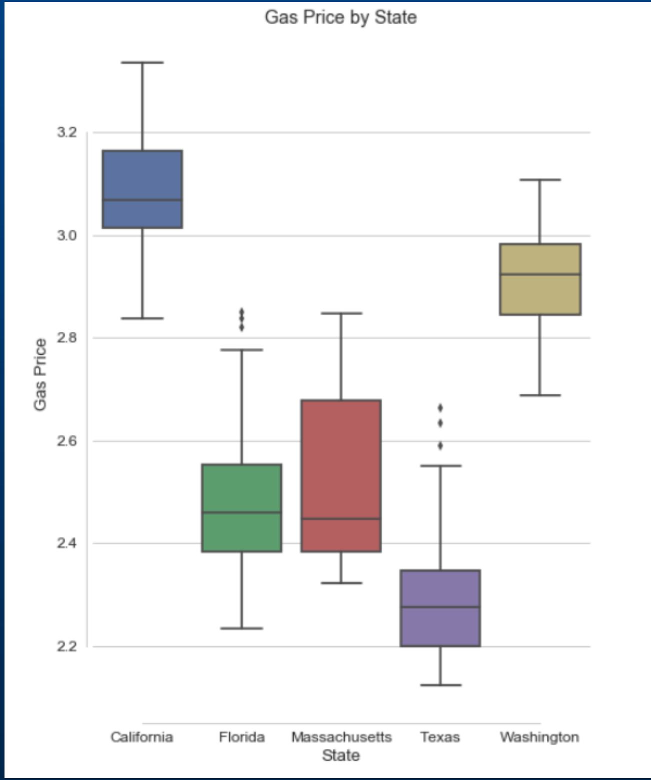


Source: https://www.eia.gov

zilliant

Gas Price:

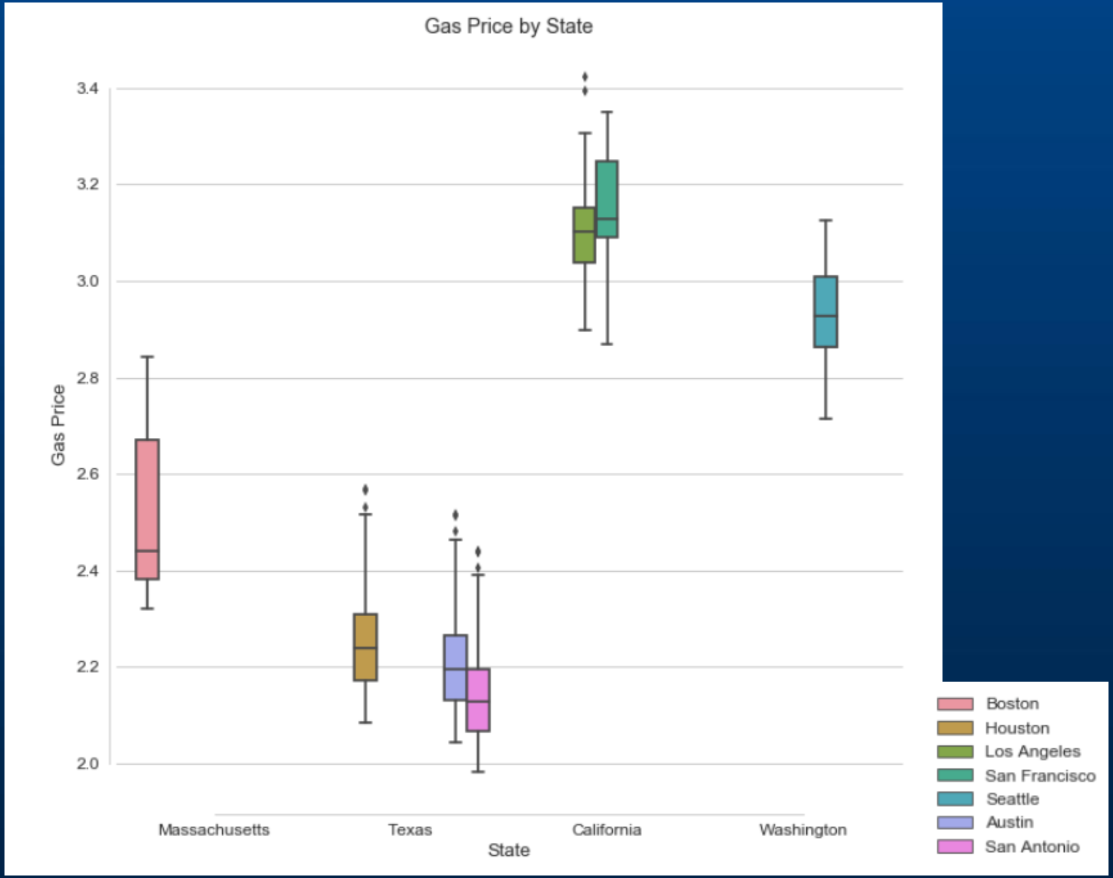




zilliant

Source: https://www.eia.gov

Gas Price:



Source: https://www.eia.gov

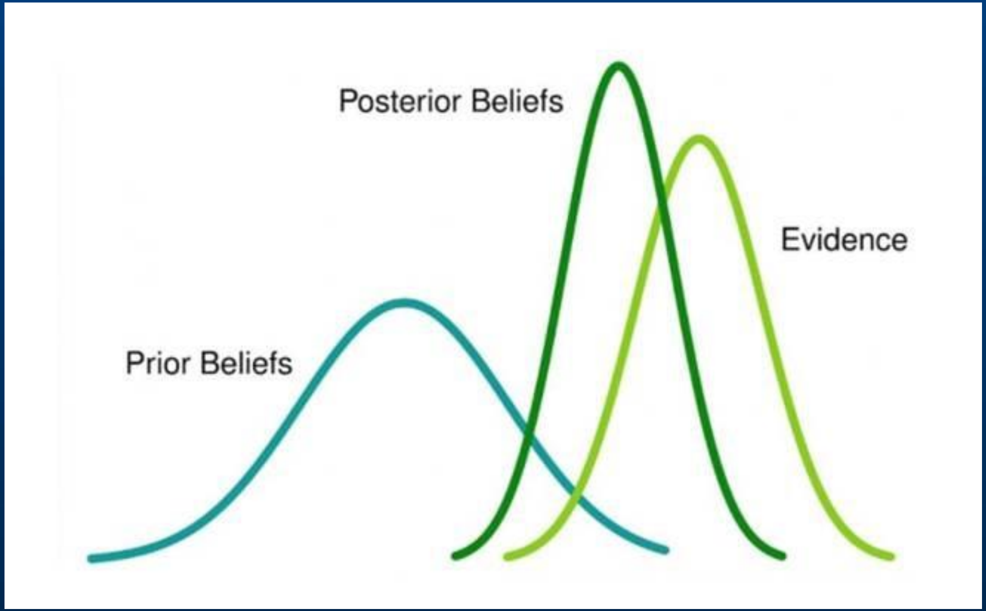
Gas Price:

San Antonio:

- Station 1: 3.5$
- Station 2: 3.8$

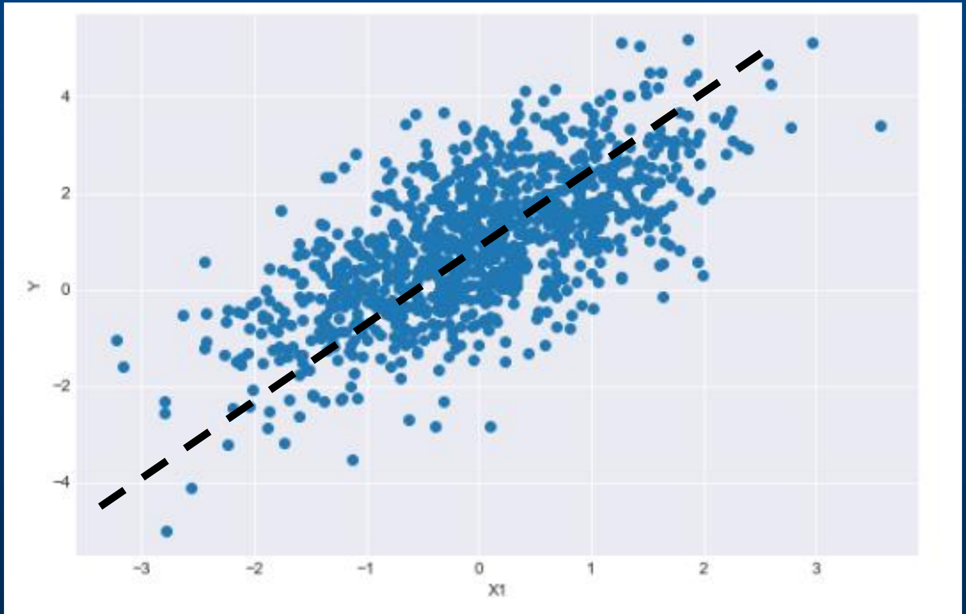
Source: https://www.eia.gov

Bayesian Models



zilliant

Bayesian Regression



Assumed relationship: $y = \alpha + \beta x$

Priors: $\alpha \sim N(0, 10)$

$\beta \sim N(1, 5)$

What other assumption(s) we have?

zilliant

Bayesian Regression

1- What is the target variable? **Y which depends on Alpha and Beta**

2- What is assumed (potential distribution)? **Normal Distribution**

3- Define priors **Alpha and/or Beta distribution**

zilliant

Bayesian Regression

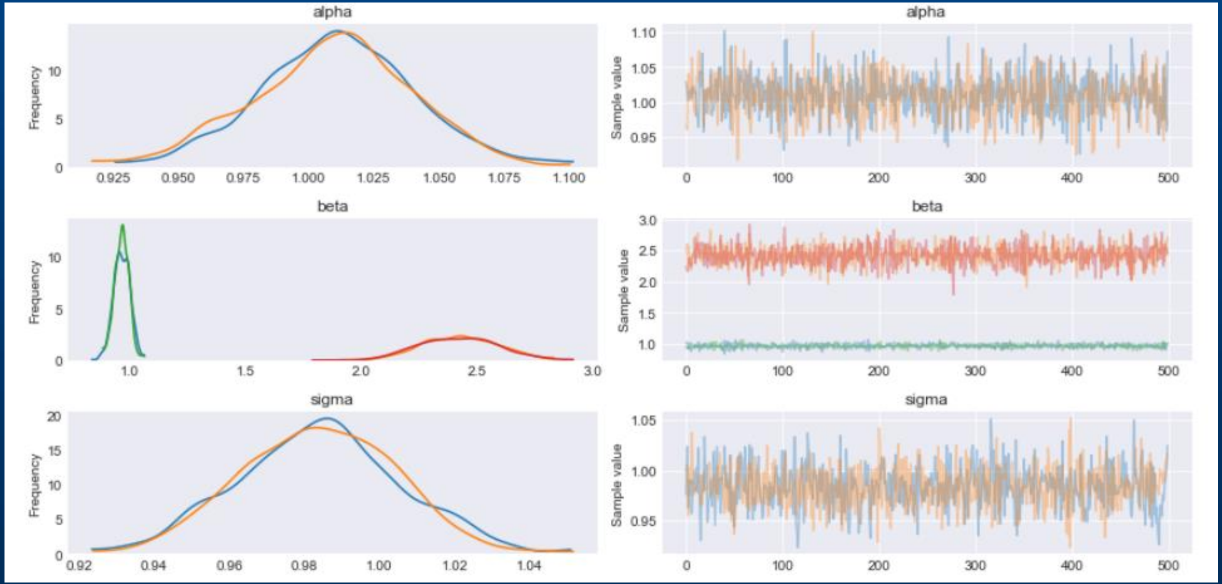
```
regression_model = pm.Model()

with regression_model:


    alpha = pm.Normal('alpha', mu=0, sd=10)
    beta = pm.Normal('beta', mu=1, sd=5)
    sigma = pm.HalfNormal('sigma', sd=1)


    mu = alpha + beta*X1


    Y_obs = pm.Normal('Y_obs', mu=mu, sd=sigma, observed=Y)
```

Create stochastic variables

Create deterministic variable

zilliant

Bayesian Regression



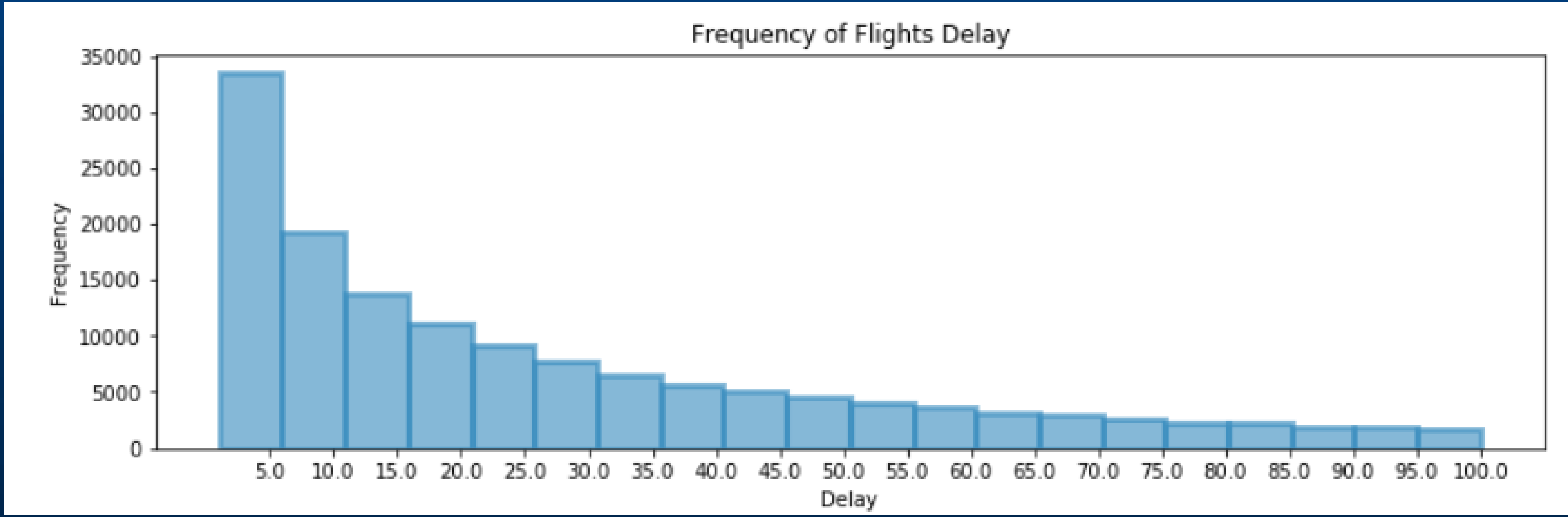
	mean	sd	mc_error	hpd_2.5	hpd_97.5	n_eff	Rhat
alpha	1.009507	0.029933	0.000790	0.950972	1.065191	1515.745188	1.000558
beta__0	0.967910	0.032986	0.000856	0.905172	1.032044	1451.511508	0.999007
beta__1	2.420182	0.164576	0.004431	2.120133	2.747157	1641.803594	0.999708
sigma	0.983615	0.021083	0.000603	0.945488	1.024915	1201.344314	0.999019

zilliant

Flight Delay Prediction

Flight Delay Prediction:

Evidence (IAH Airport)



zilliant

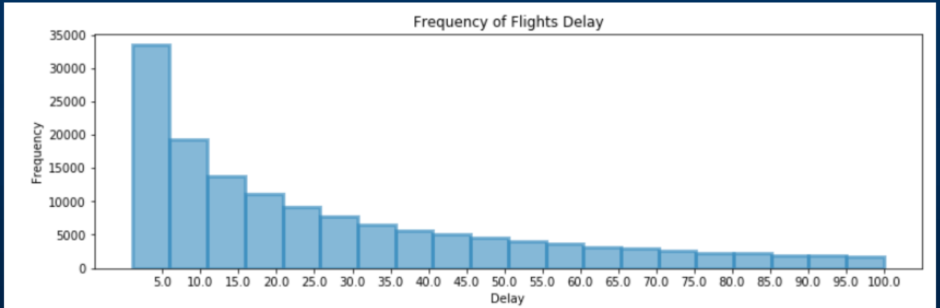
Source: http://stat-computing.org/dataexpo/2009/the-data.html

Modeling Step

1- What is the target variable?

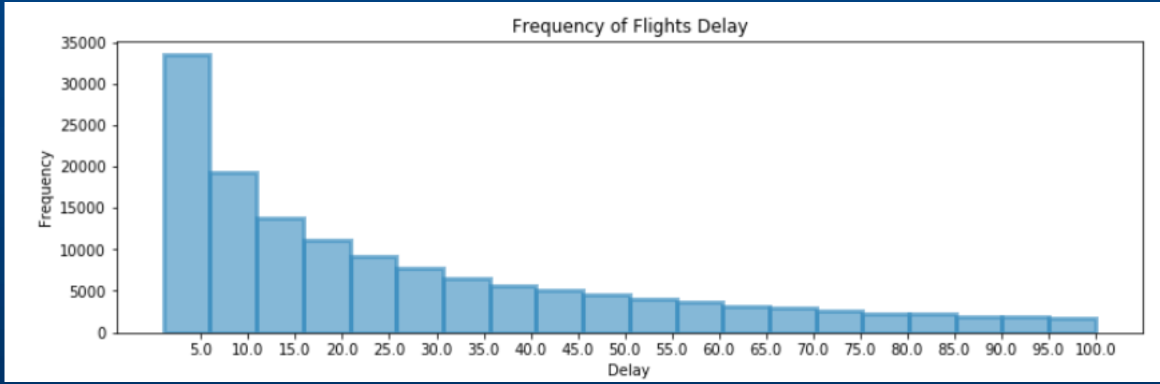
2- What is assumed (potential distribution)?

3- Define priors



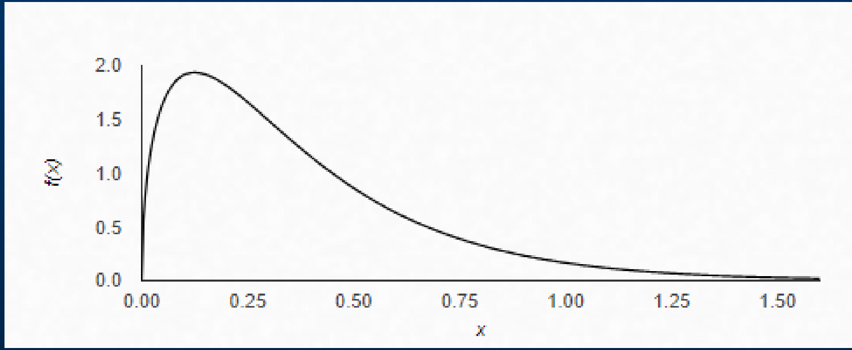
zilliant

Flight Delay Prediction:



Evidence (IAH Airport)

$$f(x;\lambda) = \lambda e^{-x\lambda}$$



Prior Belief

zilliant

Source: http://stat-computing.org/dataexpo/2009/the-data.html

Flight Delay Prediction:

```
Flight_Delay = pm.Model()

with Flight_Delay:
    # Prior Distribution

    rate = pm.Gamma('rate', 2, 2)

    # Likelihood (sampling distribution) of observations
    Y_obs = pm.Exponential('Y_obs', rate, observed=x)
```

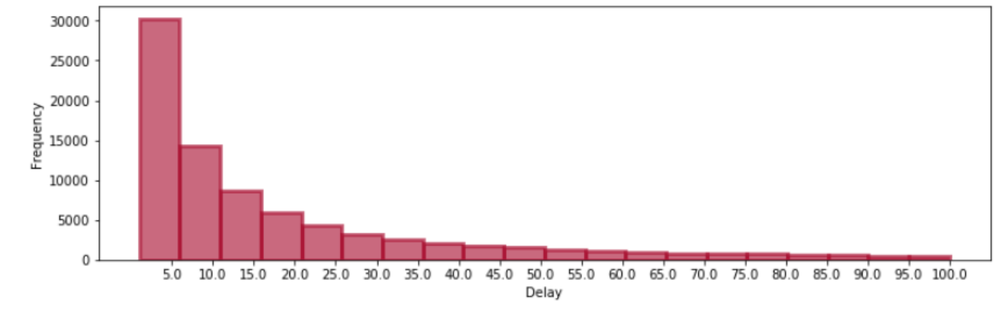
Prior Distribution of target variable

Relationship between

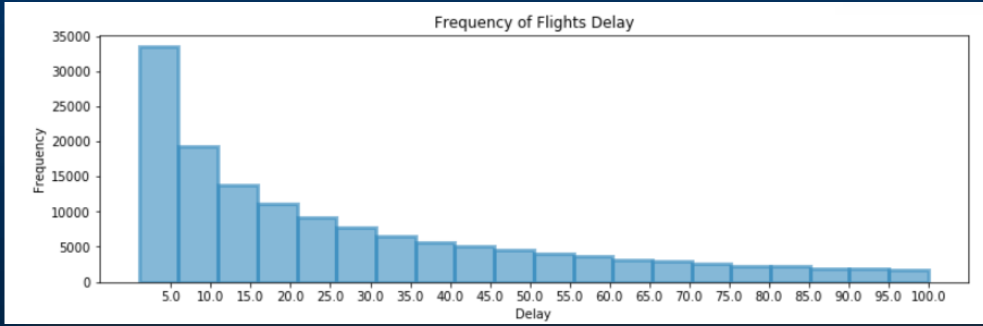
zilliant

Flight Delay Prediction:

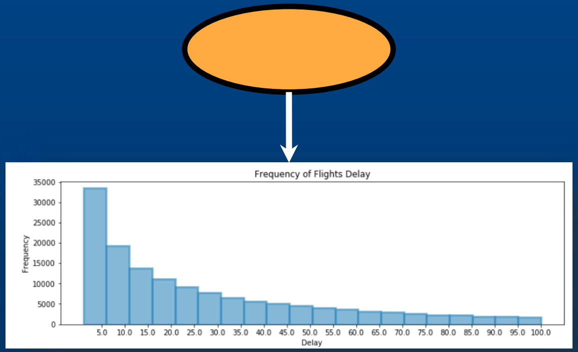
Evidence (ORD Airport)

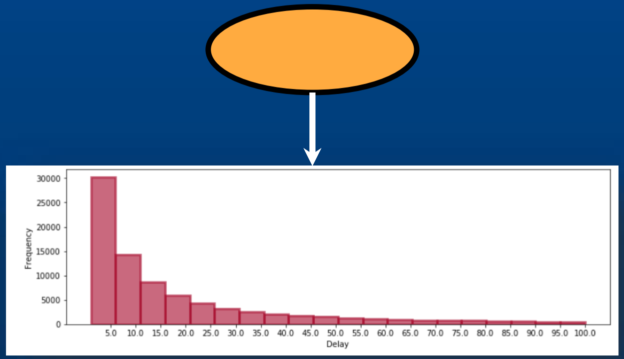


Evidence (IAH Airport)

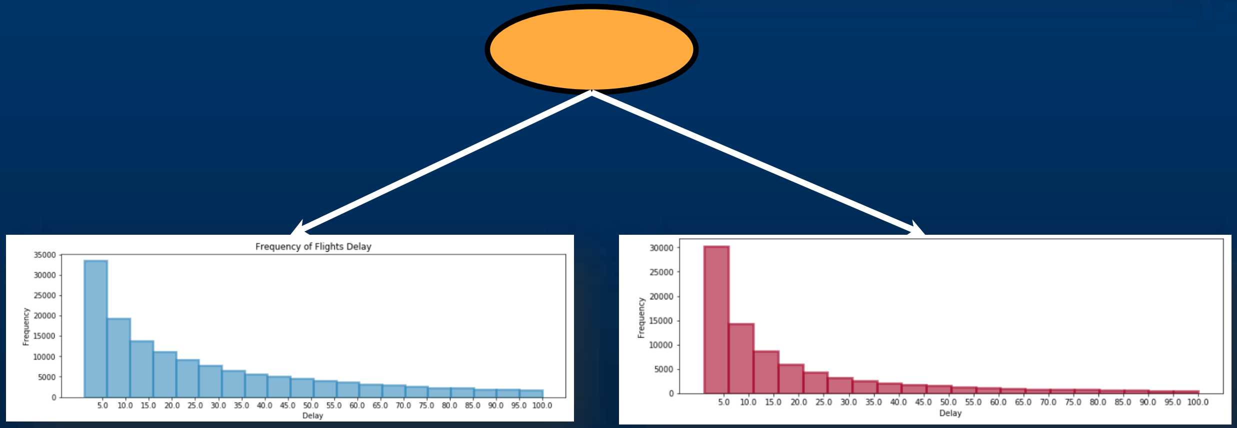


Flight Delay Prediction:





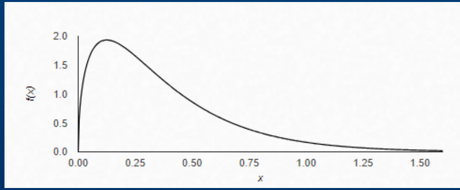
Every one has its own model!



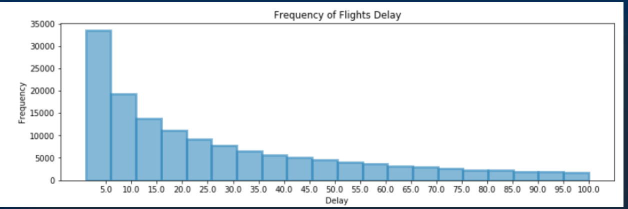
All have one model!

Flight Delay Prediction:

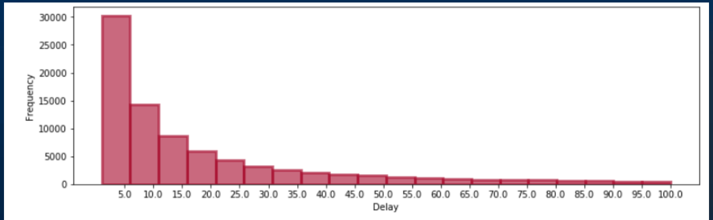
$$Gamma(\alpha_1, \beta_1)$$



$$f(x; \lambda_1) = \lambda_1 e^{-x\lambda_1}$$



$$f(x; \lambda_2) = \lambda_2 e^{-x\lambda_2}$$



zilliant

Flight Delay Prediction:

```
Flight_Delay = pm.Model()

with Flight_Delay:
    # Prior Distribution
    rate = pm.Gamma('rate', alpha, beta,shape=2)


    rate_hat=rate[airport]
    # Likelihood (sampling distribution) of observations
    Y_obs = pm.Exponential('Y_obs', rate_hat, observed=Delay)
```

Artificial Intelligence – A Definition for B2B

The capability of a machine to imitate intelligent human behavior

What if your best…

Sales Person

Pricer

Buyer

Product Manager

Financial Analyst

…could evaluate every customer relationship, every day and recommend action?

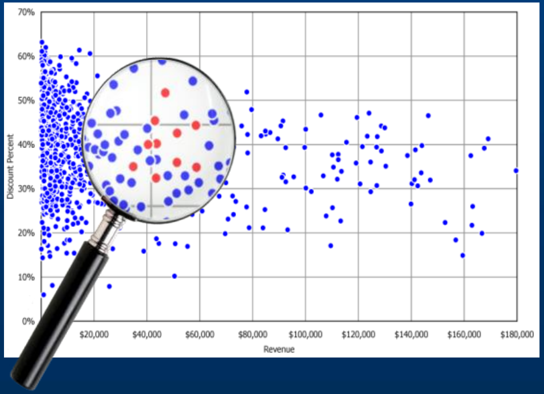
How would that impact customer lifespan, revenue and profit?

zilliant

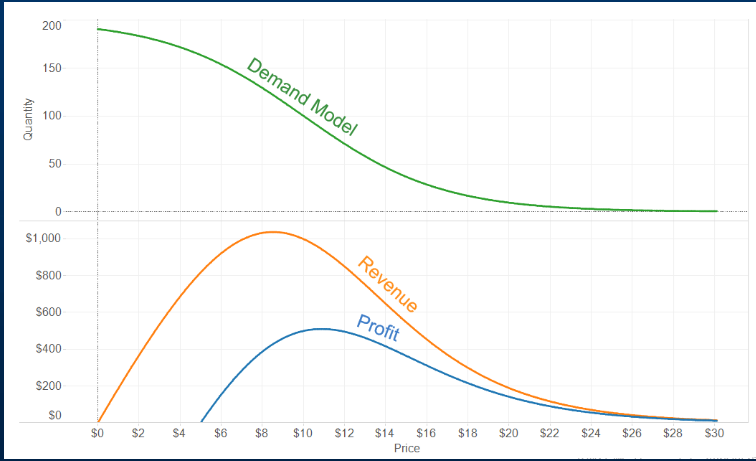
Price Optimization

Machine learning for market segmentation, price sensitivity, and price optimization.

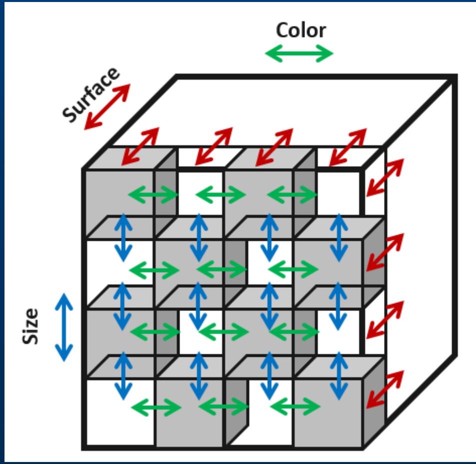
Segment Clustering



Convex Optimization

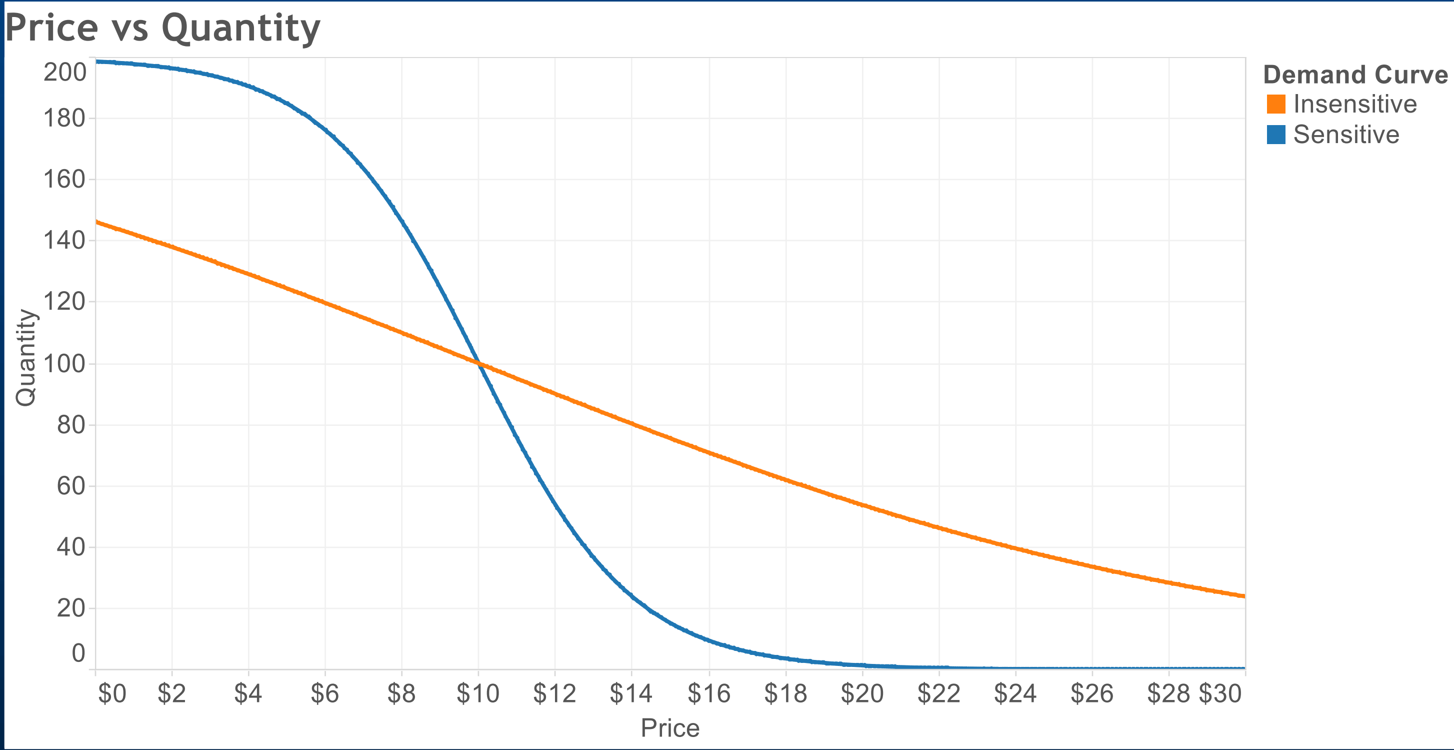


Pricing Constraints



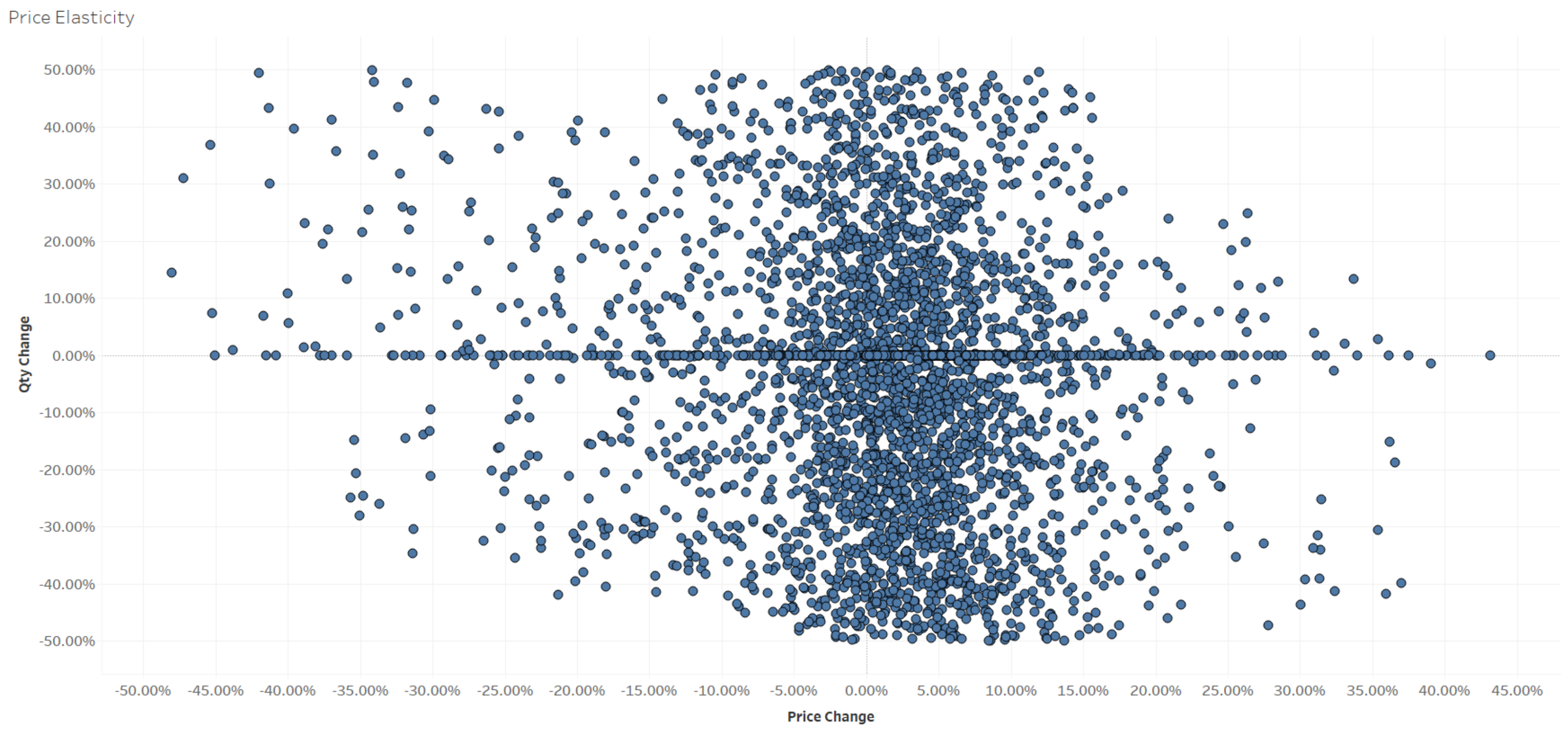
zilliant

Price Elasticity



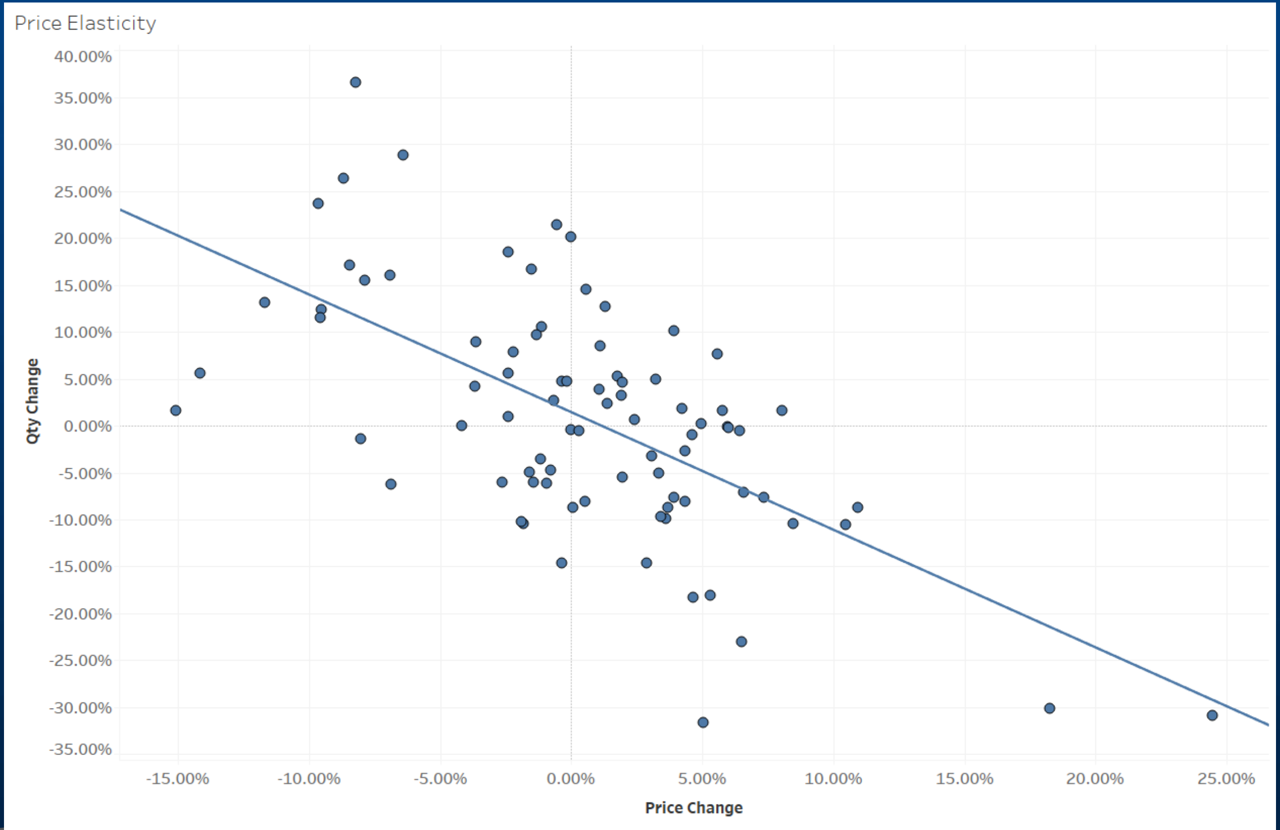
What we hope to see (Dream)

What we see (Reality!)



zilliant

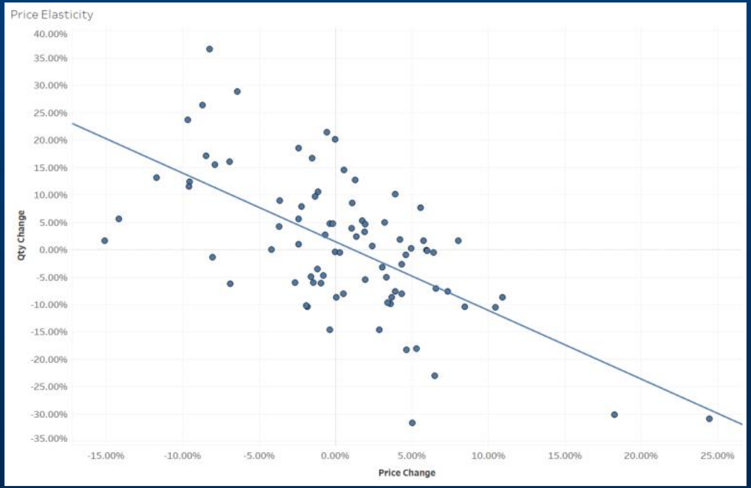
Fix it!



zilliant

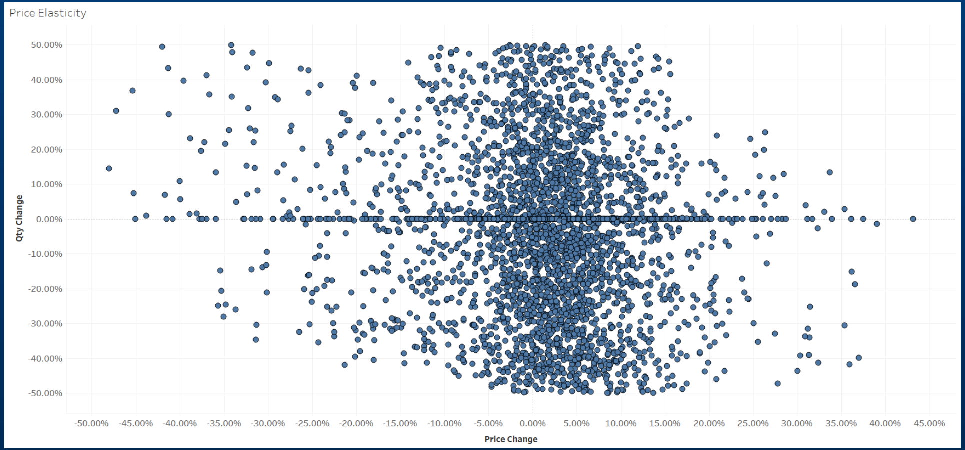
Why Hierarchical Model?

Clean Data, Less Data Points



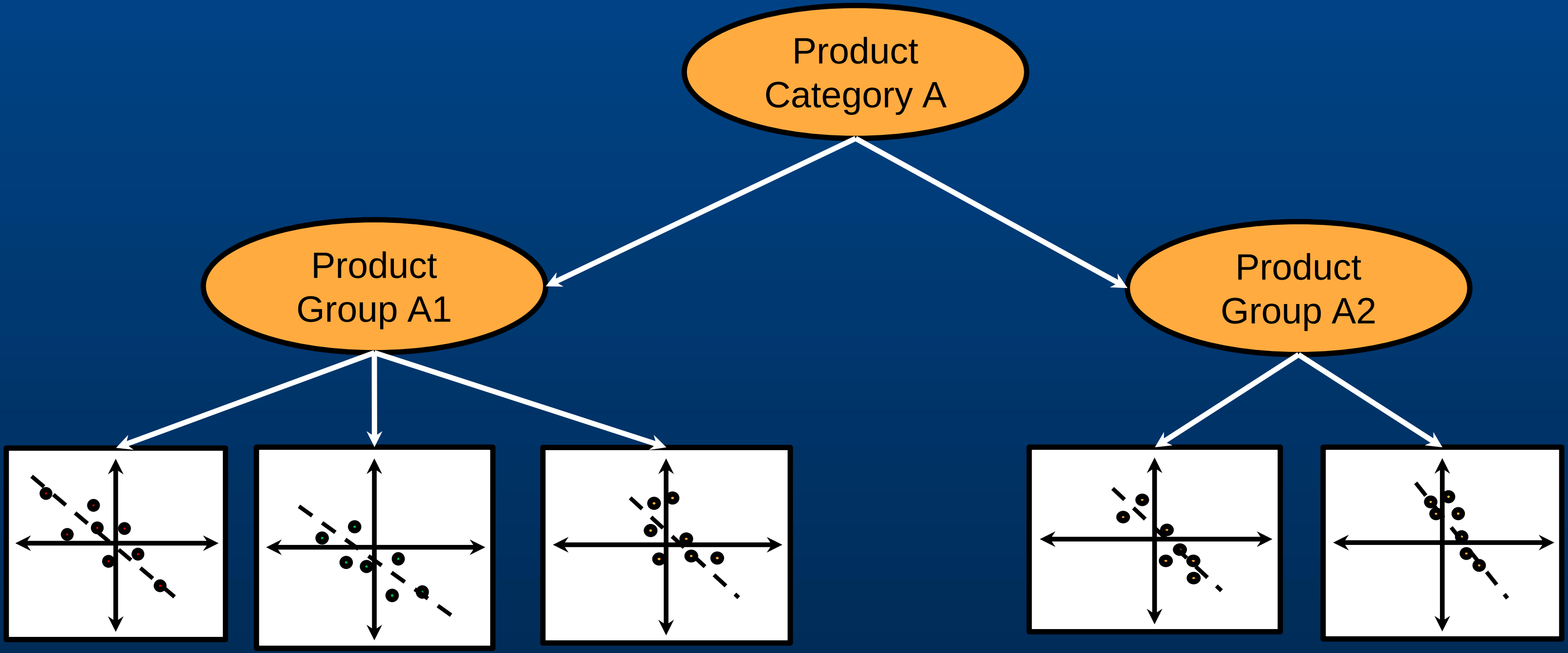
Vs.

Noisy Data, Many Data Points



zilliant

Why Hierarchical Model?



Zilliant ML Platform (Point and Click!)

Step 1- Define Hieratical Levels

Segmentation Trees: PSE

Segmentation Tree Name	PSE
Purpose	Time-Based Metrics

Segmentation Tree Levels

Use	Order	Attribute	Entity
☑	0	zProduct_Id	Transaction
☑	1	Customer_Program	Transaction
☑	2	CompetitiveRegion_Bin	Transaction
☑	3	zRegion	Transaction
☐	4	CustomerSpend_Bin	Transaction
☐	5	Customer_Loyalty	Transaction
☐	6	OrderSize_Bin	Transaction
☐	7	Price_Change_Group	Transaction
☐	8	Region	Transaction

Save Cancel

Step 2- Set Model Parameters

Regression Parameters

Enable Trace Debug	☐
Delta Days	180
Delta Lag Days	180
Max Price Delta (%)	50
Max Quantity Delta (%)	50
Transaction Threshold	20
MuA	0.0000
MuB	-1.0000
SigA	0.1000
SigB	0.2500
SigEpsHB	1.0000
SigEpsLB	0.0100
Number of Samples	10,000
Burn-in (%)	20
Calculation Time Limit (hours)	5.000
Attribute for Partitioning	1 | Customer_Program

Step 3- Run the job!

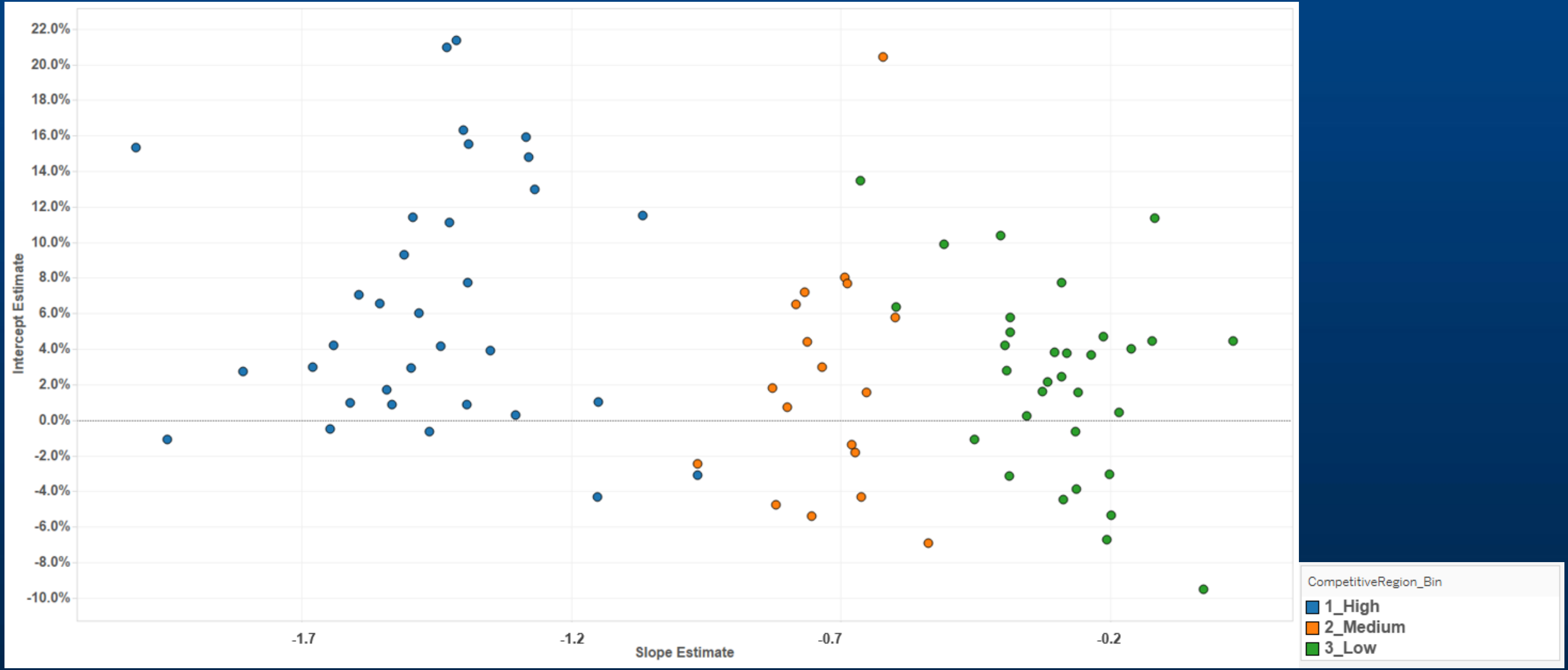
Jobs

- Compute regression

Run Job

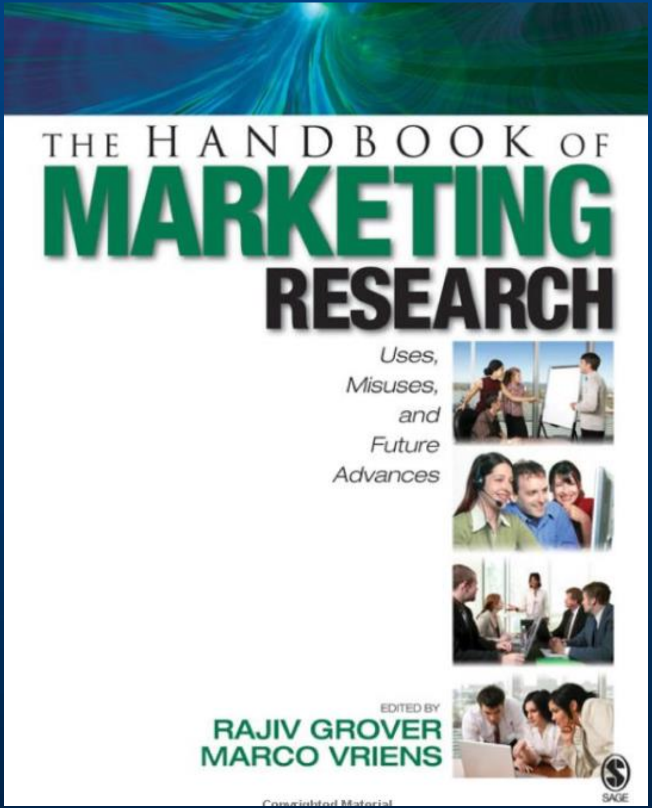
zilliant

Output



zilliant

Learn more?



20

Hierarchical Bayes Models

Greg M. Allenby
Ohio State University

Peter E. Rossi
University of Chicago

include the relationship of needs to desired attributes or wants, wants to brand beliefs and consideration sets, and consideration sets to preference orderings and choice. These extended models are often conceptualized in a hierarchical manner, where movement from one model component to the next proceeds in a logical manner. Estimation of these new integrated models is not possible without Bayesian methods.

The nature and determinants of heterogeneity have also received much attention over the past 10 years. Across dozens of studies, the distribution of heterogeneity has been shown to be better represented by a continuous, not a discrete, distribution (e.g., from a finite mixture model) of heterogeneity (Allenby, Arora, & Ginter, 1998). This has important implications for analysis connected with market segmentation, where researchers often incorrectly assert the existence of a small number of homogeneous groups. Bayesian methods are being used to identify new basis variables that point to brand preferences (Yang, Allenby, & Fennell, 2002), new ways of dealing with respondent heterogeneity in scale usage (Rossi, Gilula, & Allenby, 2001), and new ways of characterizing social networks and their impact on demand (e.g., interdependent preferences; Yang & Allenby, 2003). These developments would again not be possible without modern Bayesian methods.

In this chapter, we provide an introduction to hierarchical Bayes models and an overview of successful applications. Underlying assumptions are discussed in the next section, followed by an introduction to the computational arm of

look behind the data requires models that reflect associations of interest. Consider, for example, an analysis designed to determine the influence of price on the demand for a product or service. If the offering is available in continuous units (e.g., minutes of cell phone usage), then a regression model (see Chapter 13) can be used to measure price sensitivity using the following model:

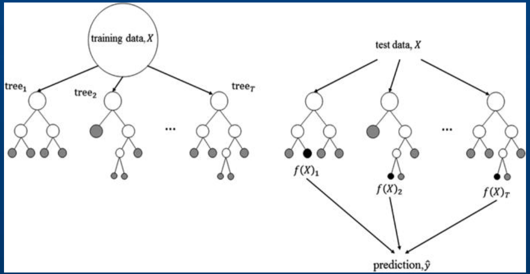
$$y_t = \beta_0 + \beta_1 \text{price}_t + \varepsilon_t \;;\; \varepsilon_t \sim \text{Normal}(0, \sigma^2) \quad (1)$$

where y_t denotes demand at time t, price_t is the price at time t, and β_0, β_1, and σ^2 are parameters to be estimated from the data. The parameters β_0 and β_1 define the expected association between price and demand. Given the price at any time, t, one can compute $\beta_0 + \beta_1\text{price}_t$ and obtain the expected demand, y_t. The parameter σ^2 is the variance of the error term ε_t and reflects the uncertainty associated with relationship. Large values of σ^2 are associated with noisy predictions, and small values of σ^2 indicate an association without much uncertainty.

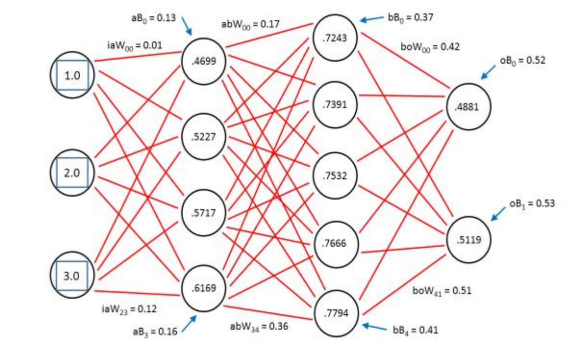
Individual-level demand, however, is rarely characterized by such a smooth, continuous association. The most frequently observed quantity of demand at the individual level is 0, and the next most frequently observed quantity is 1. Marketing data, at the individual level, are inherently discrete and noncontinuous. One approach to dealing with the discreteness of marketing data is to assume that the observed demand is a censored realization of an underlying continuous model:

The Zilliant IQ Platform Brings Best-In-Class AI to B2B

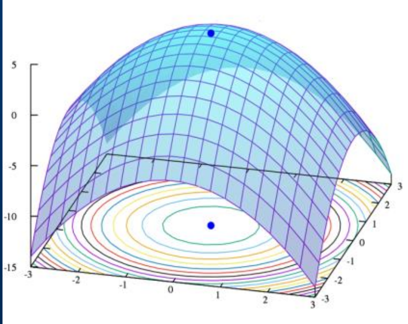
Random Forest



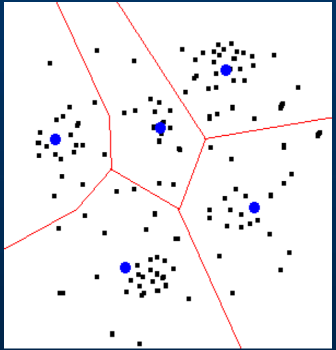
Neural Network / Deep Learning



Convex Optimization



Clustering



Association Rules

$$\text{Rule: } X \Rightarrow Y$$

$$Support = \frac{frq(X,Y)}{N}$$

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Codes and Slides:

https://github.com/AmirMK/ODSC

Question and/or Discussion:

Amir.Meimand@Zilliant.com