



IST687 Applied Data Science

Final Project Presentation



TEAM

David Forteguerre
Bradley Wuon Seok Cho
Devin Shannon
Hem Adhikari



Table of contents

I. INTRODUCTION

II. BUSINESS QUESTIONS

III. METHODOLOGY

1. Data importation, cleansing, munging, and preparation
2. NPS Calculation
3. Data descriptive analysis
4. Data modeling techniques

IV. DATA DESCRIPTIVE ANALYSIS

1. Basic descriptive analysis
2. Descriptive analysis according to NPS type

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS

1. Linear Regression
2. Association Rules
3. Support Vector Machines

VI. DATA GENERALIZATION

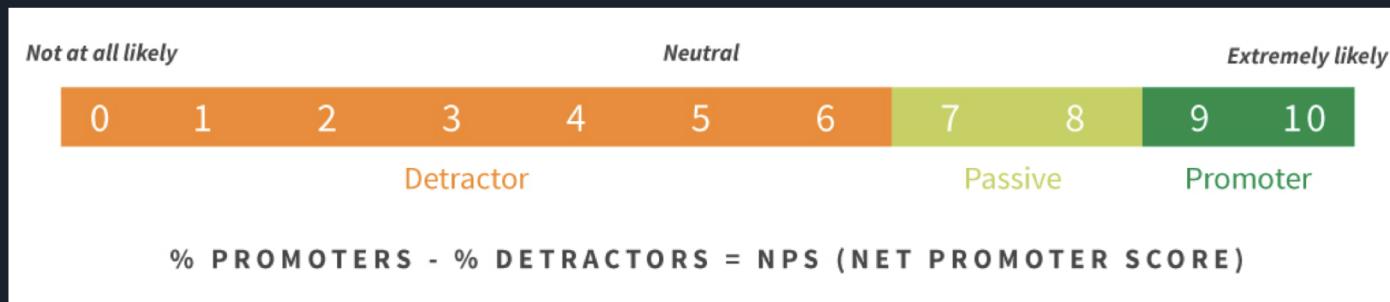
VII. DATA VALIDATION

VIII. RESULTS

1. Final Conclusions
2. Final Recommendations
3. Final Notes

I. INTRODUCTION

- We were provided with a modified version of a large dataset from the Hyatt Hotel chain and we were asked to serve as consultants
- The dataset contained customer information and survey results
- Goals:
 - identify and then answer interesting questions
 - focus on NPS (Net Promoter Score) and likelihood to recommend
 - identify the key drivers that could improve NPS
 - but no specific questions or goals were provided. The full analysis was determined by the team!





II. BUSINESS QUESTIONS

- Who are the customers? Who are the satisfied customers, and who are the unsatisfied customers?
- What factors directly affect a customer's satisfaction (more specifically, their "likelihood to recommend?")
- Which modeling techniques can be used to predict the effect on likelihood to recommend?
- How can Hyatt better target unsatisfied customers, and what steps can Hyatt management do to improve customer satisfaction?

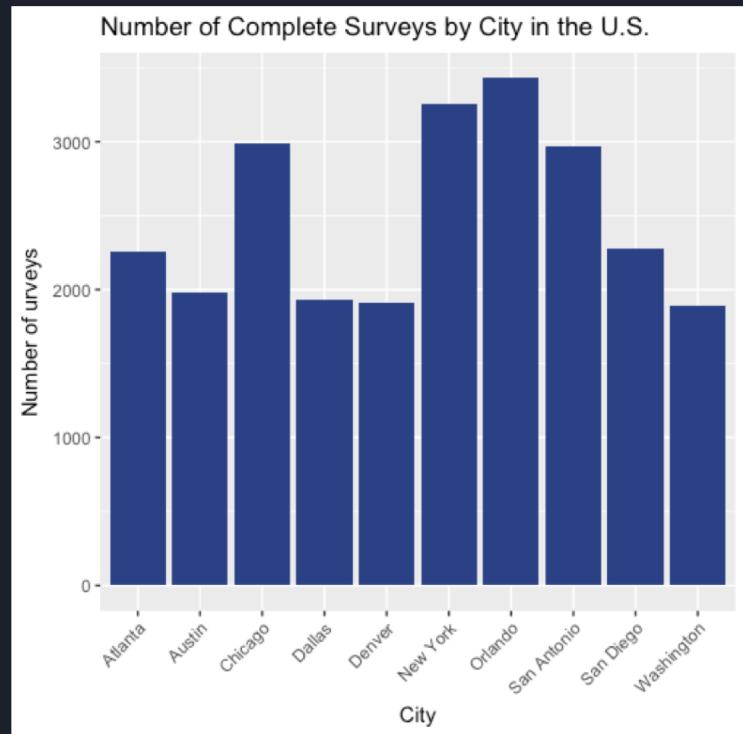
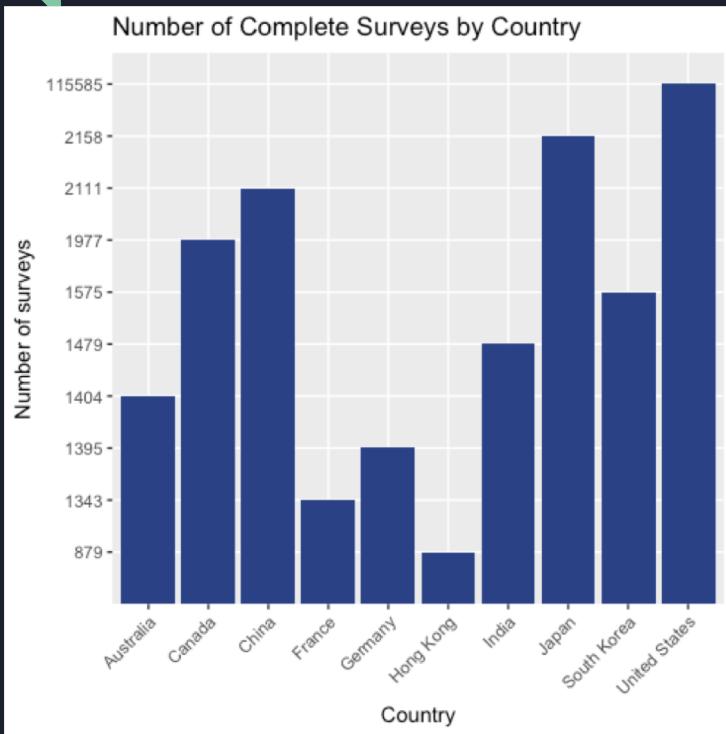
III. METHODOLOGY (1)

1. Data importation, cleansing, munging, and preparation

	Variable Name	Meaning
1.	"Country_PL"	Country in which the hotel is located
2.	"City_PL"	City in which the hotel is located
3.	"Spirit_PL"	Unique <u>hotel</u> identifier (5-letter code)
4.	"Hotel Name-Long_PL"	Full hotel name
5.	"NPS_Type"	Indicates if the guest's HySat responses mark them as a promoter, a passive, or a detractor
6.	"Likelihood_Recommend_H"	Likelihood to recommend metric; value on a 1 to 10 scale
7.	"POV_CODE_C"	Purpose of visit
8.	"MEMBER_STATUS_R"	Tier of the GP program that the member belongs
9.	"LENGTH_OF_STAY_C"	Length of stay
10.	"GUEST_COUNTRY_R"	The country to which the actual guest belongs to. Different from Country_Code
11.	"Gender_H"	Guest's gender
12.	"Age_Range_H"	Guest's age range
13.	"Overall_Sat_H"	Overall satisfaction metric; value on a 1 to 10 scale
14.	"Guest_Room_H"	Guest room satisfaction metric; value on a 1 to 10 scale
15.	"Tranquility_H"	Tranquility metric; value on a 1 to 10 scale
16.	"Condition_Hotel_H"	Condition of hotel metric; value on a 1 to 10 scale
17.	"Customer_SVC_H"	Quality of customer service metric; value on a 1 to 10 scale
18.	"Staff_Cared_H"	Staff cared metric; value on a 1 to 10 scale
19.	"Internet_Sat_H"	Internet satisfaction metric; value on a 1 to 10 scale
20.	"Check_In_H"	Quality of the check in process metric; value on a 1 to 10 scale
21.	"F&B_Overall_Experience_H"	Overall F&B experience metric; value on a 1 to 10 scale

III. METHODOLOGY (2)

1. Data importation, cleansing, munging, and preparation



III. METHODOLOGY (3)

1. Data importation, cleansing, munging, and preparation

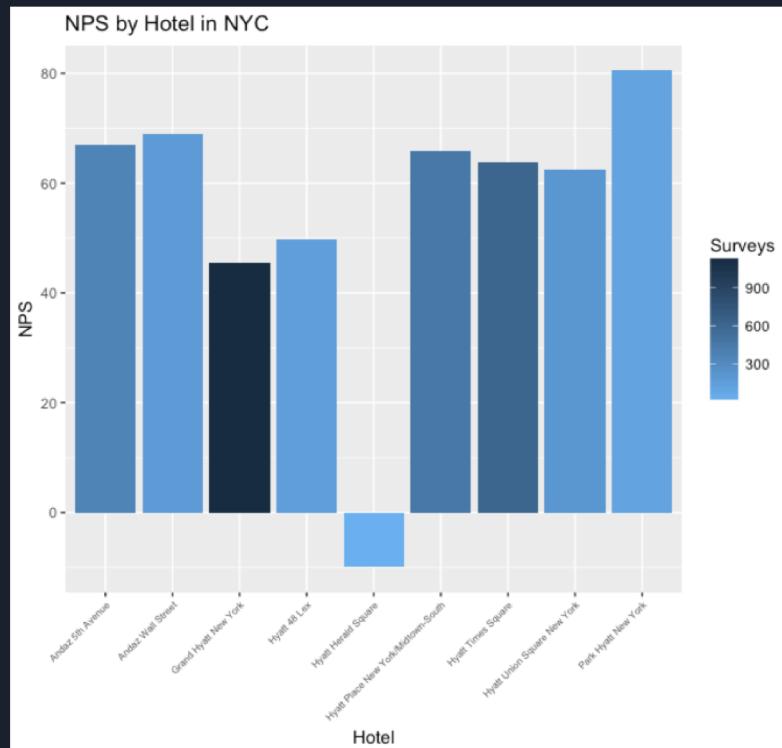


```
> HotelServicesSummary
  All.Suites_PL Bell.Staff_PL Boutique_PL Business.Center_PL Casino_PL Conference_PL Convention_PL Dry.Cleaning_PL Elevators_PL
N      6          1          7          2          7          7          6          0          0
Y      1          3          0          2          0          0          1          4          4
  Fitness.Center_PL Fitness.Trainer_PL Golf_PL Indoor.Corridors_PL Laundry_PL Limo.Service_PL Mini.Bar_PL Pool.Indoor_PL Pool.Outdoor_PL
N      0          3          7          0          1          0          1          4          4
Y      4          1          0          7          3          4          3          0          0
  Regency.Grand.Club_PL Resort_PL Restaurant_PL Self.Parking_PL Shuttle.Service_PL Ski_PL Spa_PL Spa.services.in.fitness.center_PL
N      3          7          1          4          4          7          6          3
Y      1          0          6          0          0          0          1          0
  Spa.online.booking_PL Spa.F.B.offering_PL Valet.Parking_PL
N      1          1          1
Y      0          0          3
```

III. METHODOLOGY (3)

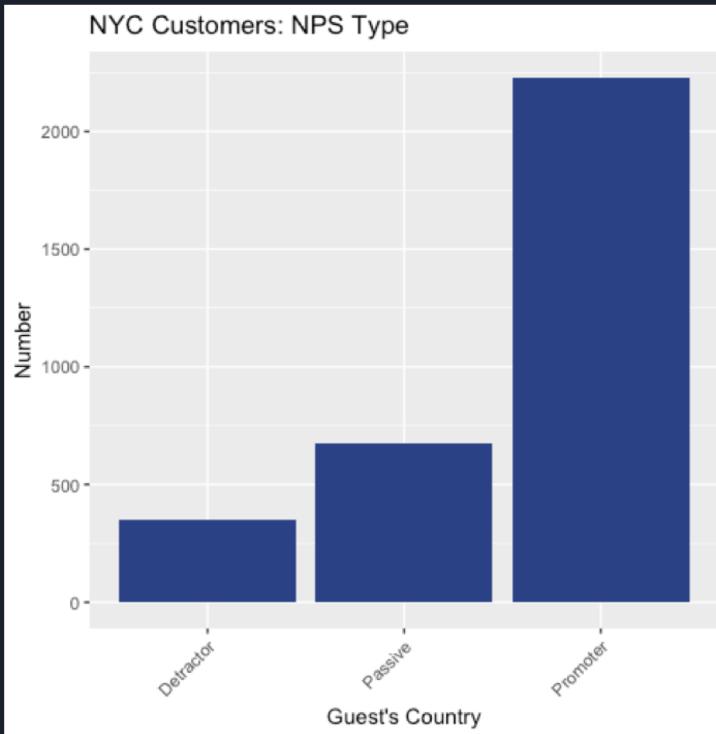
2. NPS Calculation

	HotelName	Surveys	NPS
1	Grand Hyatt New York	1130	45.48673
2	Hyatt Times Square	607	63.75618
3	Hyatt Place New York/Midtown-South	462	65.80087
4	Andaz 5th Avenue	365	66.84932
5	Hyatt Union Square New York	219	62.55708
6	Andaz Wall Street	177	68.92655
7	Hyatt 48 Lex	155	49.67742
8	Park Hyatt New York	118	80.50847
9	Hyatt Herald Square	20	-10.00000



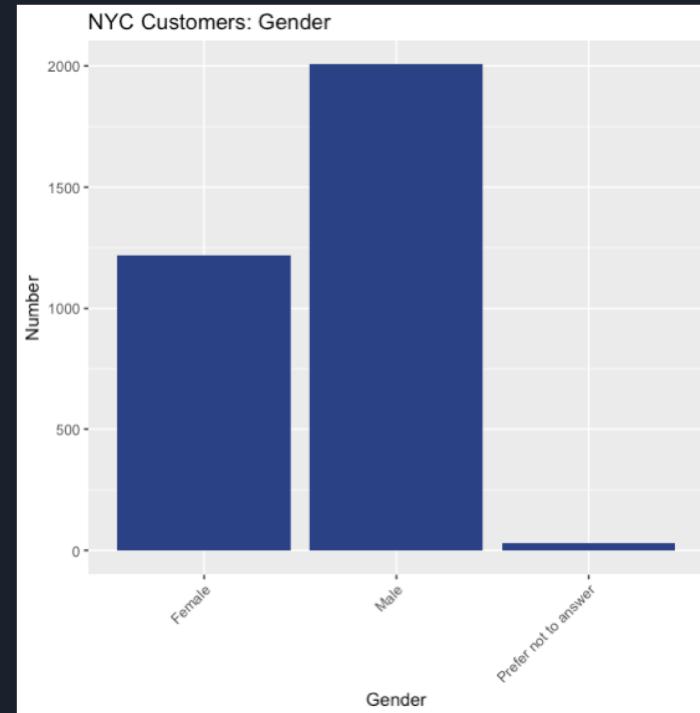
IV. DATA DESCRIPTIVE ANALYSIS (1)

1. Basic descriptive analysis



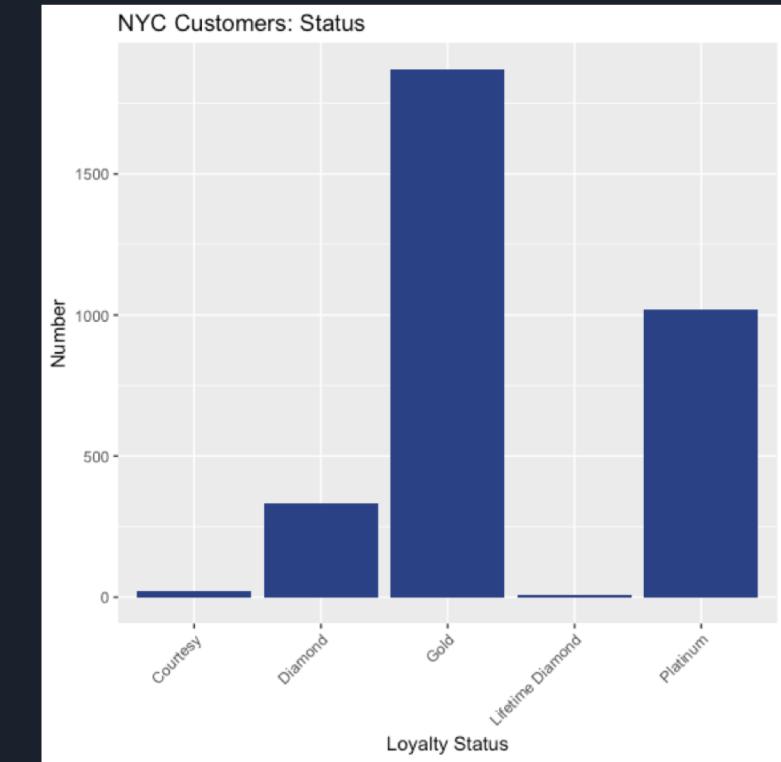
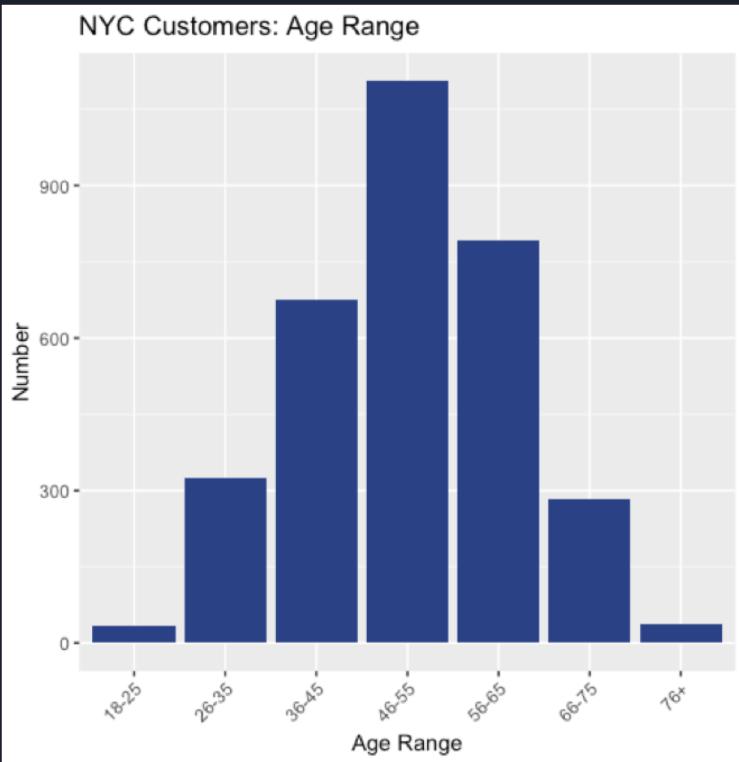
IV. DATA DESCRIPTIVE ANALYSIS (2)

1. Basic descriptive analysis



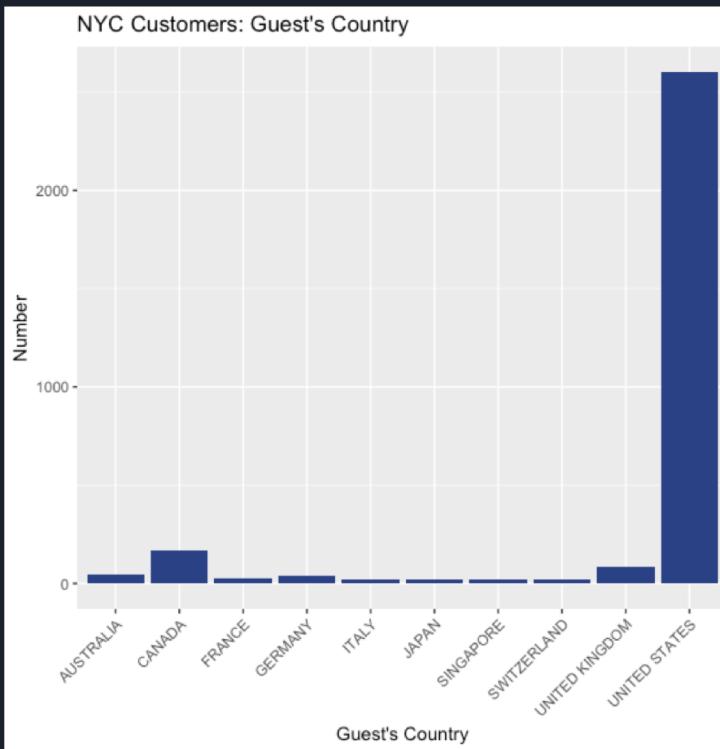
IV. DATA DESCRIPTIVE ANALYSIS (3)

1. Basic descriptive analysis



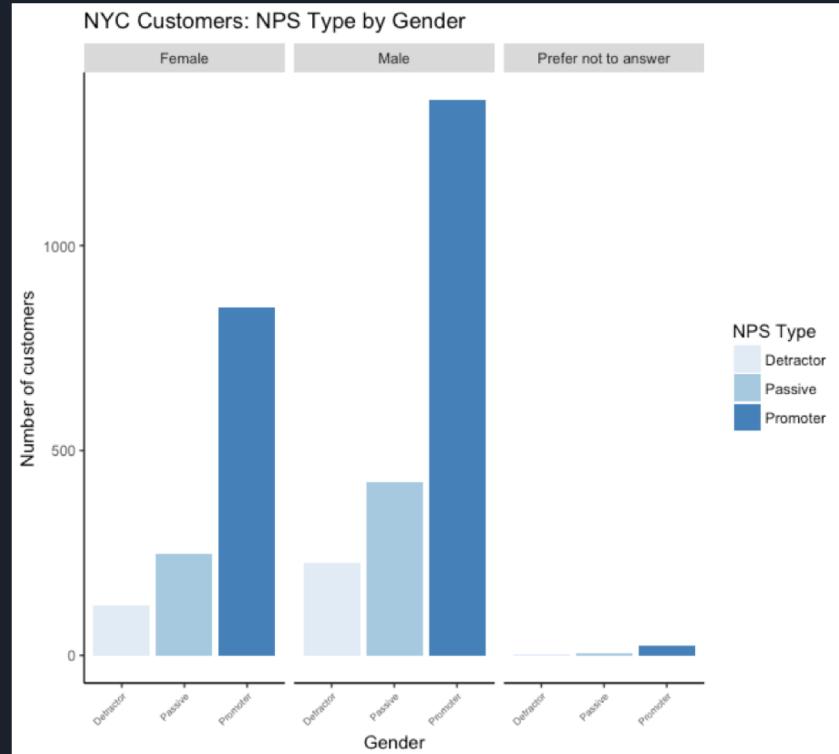
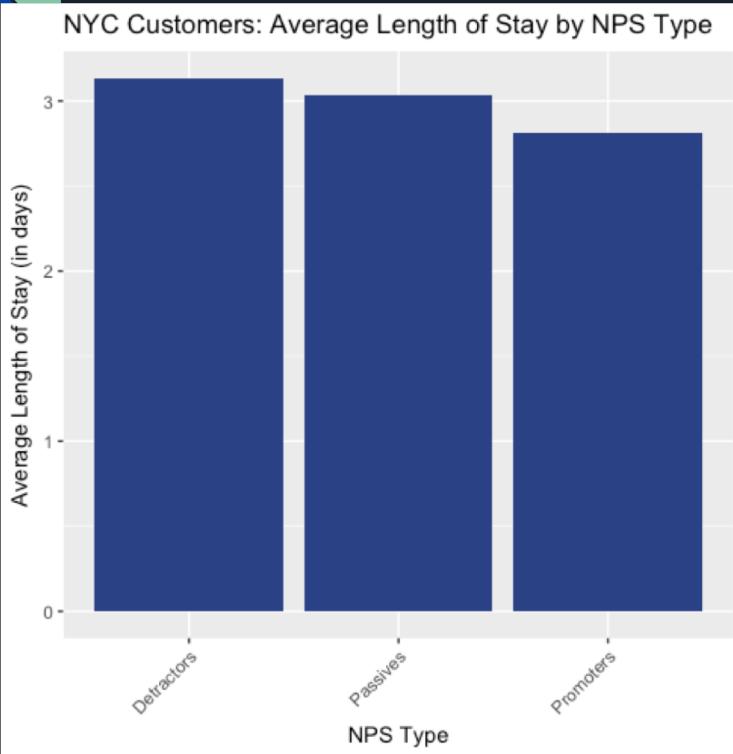
IV. DATA DESCRIPTIVE ANALYSIS (4)

1. Basic descriptive analysis



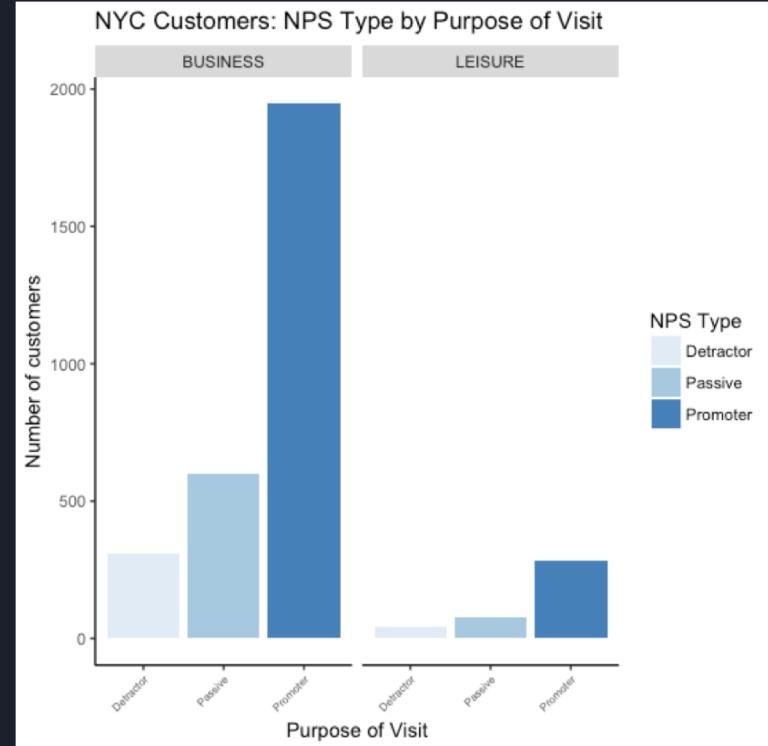
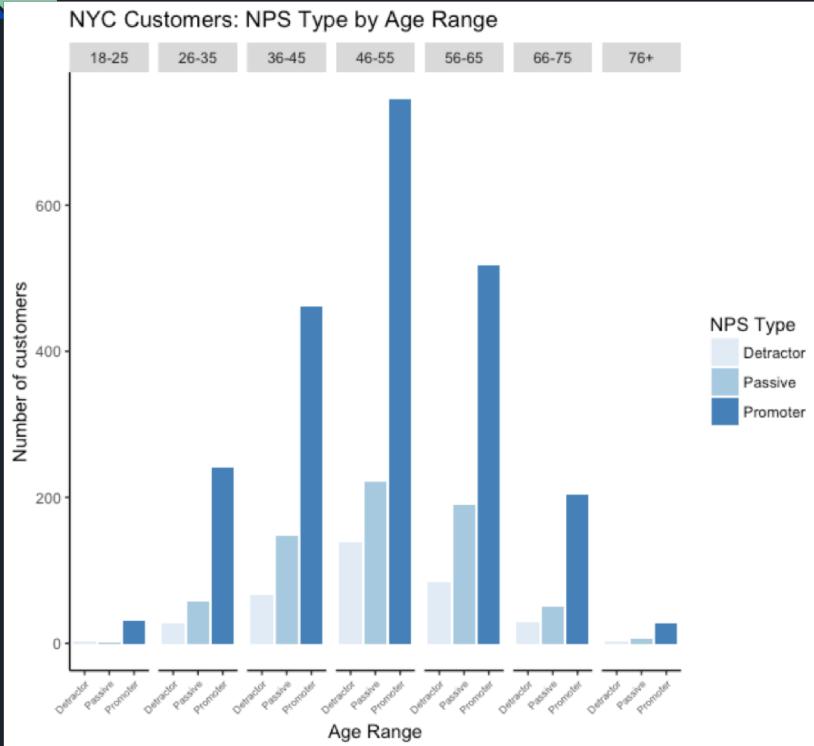
IV. DATA DESCRIPTIVE ANALYSIS (5)

2. Descriptive analysis according to NPS type



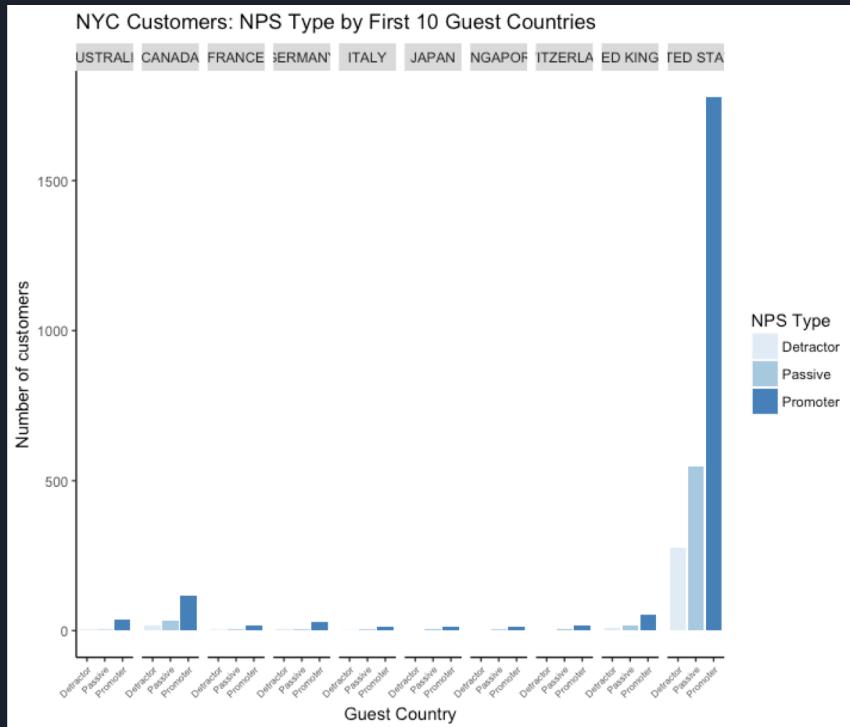
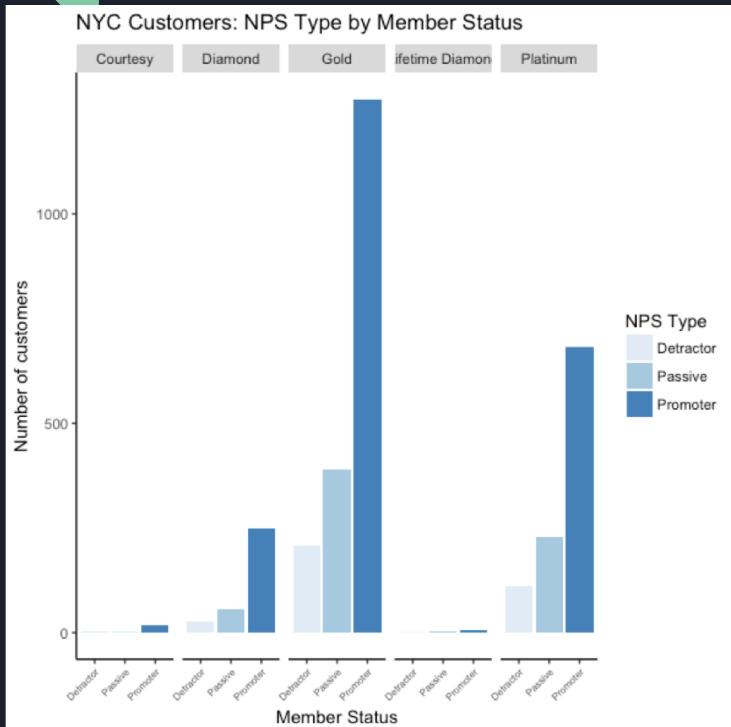
IV. DATA DESCRIPTIVE ANALYSIS (6)

2. Descriptive analysis according to NPS type



IV. DATA DESCRIPTIVE ANALYSIS (7)

2. Descriptive analysis according to NPS type





IV. DATA DESCRIPTIVE ANALYSIS (7)

Observations:

1. The NPS distribution seems to be proportional overall among the demographics.
2. NPS Type, seems not to be influenced by any demographics in particular, rather it reinforces our assumption that NPS Type is likely connected to the amenities and services provided by the establishments.

First recommendations: *(for marketing purposes)*

1. NYC is an international tourist hub, as such it is disappointing that the majority of the customers are U.S.-based travellers. The Hyatt Hotel Corporation should try to attract more international travellers.
2. Hyatt Hotel Corporation should consider a wider variety of demographics:
 - a. Female travelers
 - b. Leisure travelers
 - c. Younger and older travelers besides (40-60)

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Modeling

Questions:

- Do certain factors influence a customer's decision to recommend?
- In what way?
- How do we express it?
- Can we quantify it?
- What suggestions would we make?



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Tasks:

1. Establish the null hypothesis:
 - a. assume that [x]variable(s) have no effect at all on the dependent [Y]variable (in this case, "Likelihood to Recommend" used as the dependant variable).
2. Seek to either disprove or confirm the null hypothesis
3. Refine the dataset
 - a. Identify useful variables for the analysis
 - b. Eliminate unneeded variables
4. Run a variety of regression analyses
 - a. Analyze the statistical information
 - b. Plot the results
5. Eliminate unnecessary/inconsequential variables
6. Rerun the analyses
7. Analyze the results
8. Determine the most influential components of the customer experience
9. Make recommendations

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Variables (and definitions) used

Dependant Variable

- Likelihood_Recommend_H - Likelihood to recommend metric; value on a 1 to 10 scale

Independant:

- Check_In_H - Quality of the check in process metric; value on a 1 to 10 scale [*Eventually Removed*]
- Condition_Hotel_H - Condition of hotel metric; value on a 1 to 10 scale
- Customer_SVC_H - Quality of customer service metric; value on a 1 to 10 scale
- F&B_Overall_Experience_H - Overall F&B experience metric; value on a 1 to 10 scale
- Guest_Room_H - Guest room satisfaction metric; value on a 1 to 10 scale
- Internet_Sat_H - Internet satisfaction metric; value on a 1 to 10 scale [*Eventually Removed*]
- Overall_Sat_H - Overall satisfaction metric; value on a 1 to 10 scale [*Eventually Removed*]
- Staff_Cared_H - Staff cared metric; value on a 1 to 10 scale
- Tranquility_H - Tranquility metric; value on a 1 to 10 scale



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

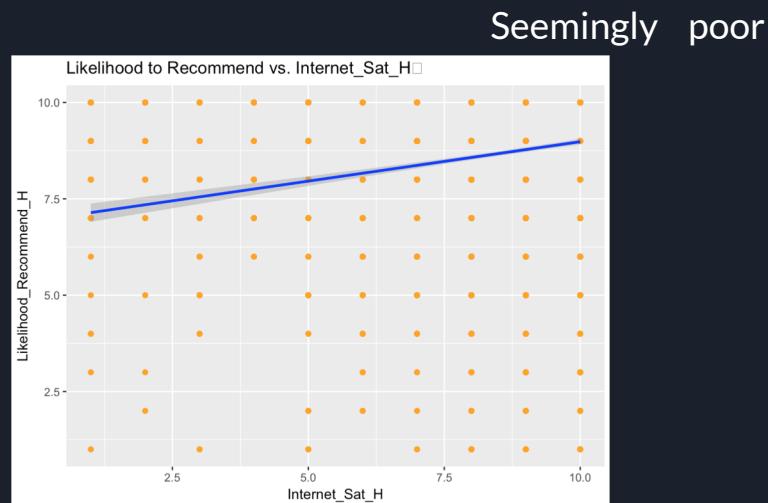
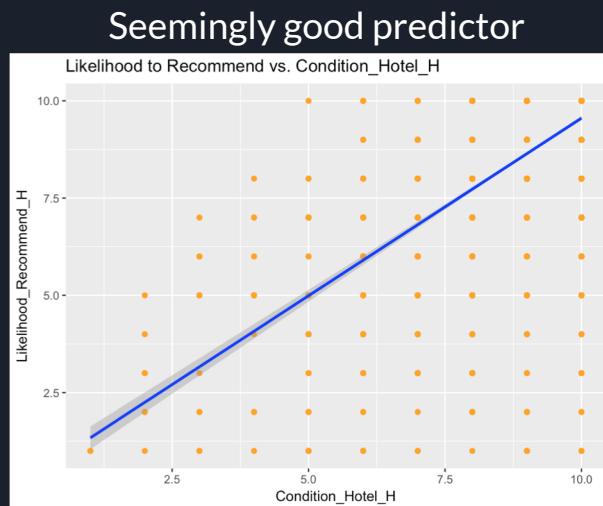
Steps:

- 1) Perform exploratory analysis using a **simple linear regression** (each individual variable againsts the dependant variable).
 - a) Analyze statistical information
 - b) Plot the results
 - c) Observe and make note of variables with potential influential value
- 1) Perform exploratory analysis using **multiple linear regression** (different combinations of variable againsts the dependant variable).
 - a) Analyze statistical information
 - b) Plot the results
 - c) Observe and make note of variables with potential influential value
- 1) Perform predictive analysis using **multiple linear regression & backward elimination** (different combinations of variable againsts the dependant variable).
 - a) Analyze statistical information
 - b) Plot the results
 - c) Observe and make note of variables with potential influential value

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Simple Linear Regression Output:



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Multiple Linear Regression Output:

Seemingly good combo of predictors



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Backward Elimination (most parsimonious predictive model) Output:

Best combo of predictors

```
Start: AIC=170.22
Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H +
Customer_SVC_H + Staff_Cared_H + F.B_Overall_Experience_H

          Df Sum of Sq    RSS   AIC
<none>                 3413.0 170.22
- F.B_Overall_Experience_H 1     19.39 3432.4 186.65
- Staff_Cared_H            1     47.03 3460.1 212.74
- Tranquility_H           1     55.80 3468.8 220.97
- Condition_Hotel_H        1     68.24 3481.3 232.62
- Customer_SVC_H          1     277.39 3690.4 422.41
- Guest_Room_H             1     332.85 3745.9 470.93

Call:
lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H +
Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + F.B_Overall_Experience_H,
data = dataset)

Coefficients:
              (Intercept)          Guest_Room_H          Tranquility_H
                           -1.37218                   0.29947                   0.09873
                           0.17368
              Customer_SVC_H          Staff_Cared_H          F.B_Overall_Experience_H
                           0.37917                   0.13692                   0.04791
```

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (1)

1. Linear Regression

Recommendations and suggestions

1. Focus on all of the variables as each have an effect on the Likelihood to recommend
 - a. Put most emphasis should be put specifically on the top two variables:
 - i. Customer_SVC_H
 - ii. Guest_Room_H
 - b. Put least emphasis on bottom two variables (contain much less influence):
 - i. Tranquility_H
 - ii. F.B_Overall_Experience_H
1. Cultivate and maintain a strong customer-oriented experience

Case and point:

If staff were able to increase the ‘Customer_SVC_H’ score alone (has ~ 40% influence on likelihood to recommend), by more than two points, you will see an increase of one or more points in the Likelihood to recommend. Even more, a combination of increases in multiple variables will allow you to transform a detractor into a passive/promoter and passive into a promoter!

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (2)

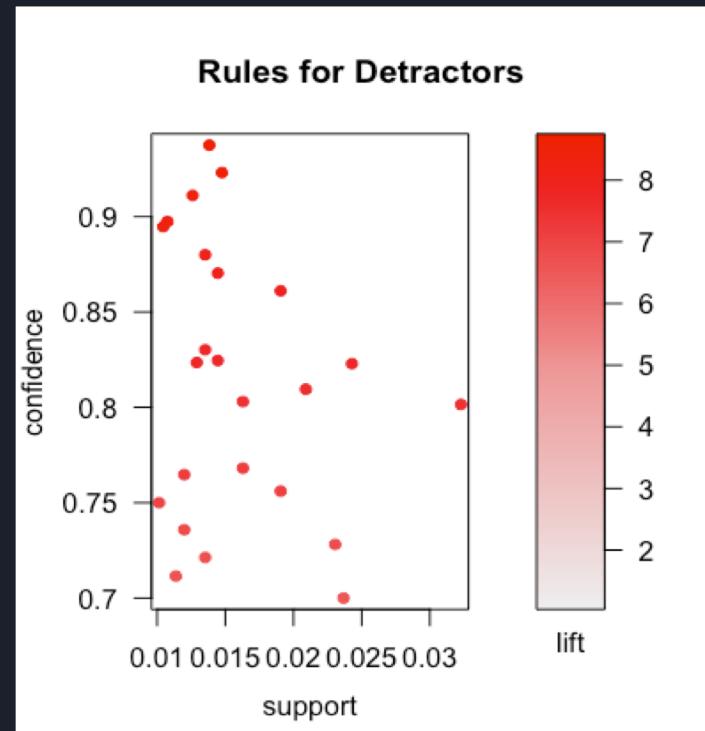
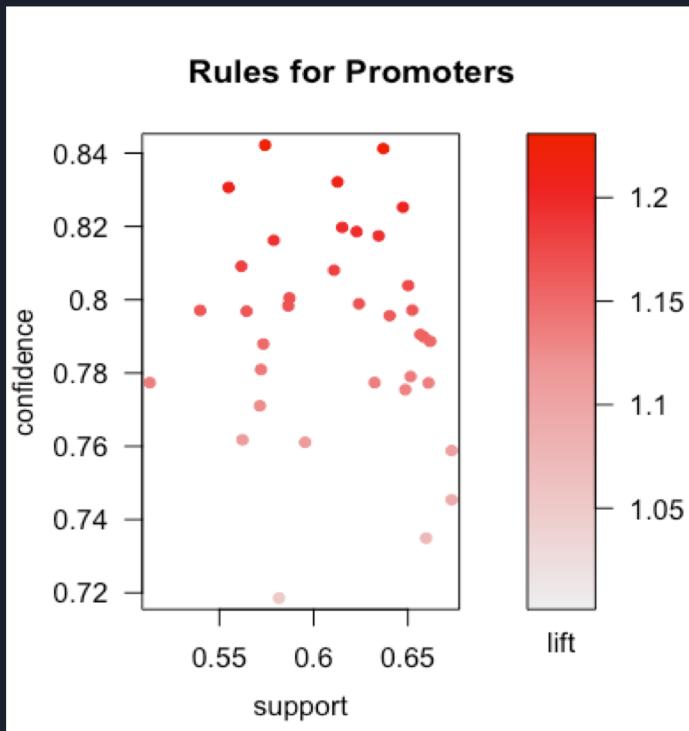
2. Association Rules

Columns within the New York City dataset

1. NPS_Type
2. Guest_Room_H
3. Tranquility_H
4. Condition_Hotel_H
5. Customer_SVC_H
6. Staff_Cared_H
7. Internet_Sat_H
8. Check_In_H
9. F.B_Overall_Experience_H

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (2)

2. Association Rules



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (2)

2. Association Rules

Rules for Promoters:

> inspect(RulesPromoters)	lhs	rhs	support	confidence	lift	count
[1]	{Tranquility_H=Tranquility_H_HIGH, F.B_Overall_Experience_H=F.B_Overall_Experience_H_HIGH} => {NPS_Type=Promoter}	0.5548724	0.8306489	1.212792	1805	
[2]	{Guest_Room_H=Guest_Room_H_HIGH, F.B_Overall_Experience_H=F.B_Overall_Experience_H_HIGH} => {NPS_Type=Promoter}	0.5742392	0.8422002	1.229658	1868	
[3]	{Staff_Cared_H=Staff_Cared_H_HIGH, F.B_Overall_Experience_H=F.B_Overall_Experience_H_HIGH} => {NPS_Type=Promoter}	0.5788503	0.8162115	1.191713	1883	
[4]	{Customer_SVC_H=Customer_SVC_H_HIGH, F.B_Overall_Experience_H=F.B_Overall_Experience_H_HIGH} => {NPS_Type=Promoter}	0.5871503	0.8005029	1.168777	1910	
[5]	{Guest_Room_H=Guest_Room_H_HIGH, Internet_Sat_H=Internet_Sat_H_HIGH} => {NPS_Type=Promoter}	0.5616354	0.8091231	1.181363	1827	
[6]	{Guest_Room_H=Guest_Room_H_HIGH, Tranquility_H=Tranquility_H_HIGH} => {NPS_Type=Promoter}	0.6151245	0.8197460	1.196873	2001	
[7]	{Tranquility_H=Tranquility_H_HIGH, Staff_Cared_H=Staff_Cared_H_HIGH} => {NPS_Type=Promoter}	0.6126652	0.8321503	1.214984	1993	
[8]	{Tranquility_H=Tranquility_H_HIGH, Customer_SVC_H=Customer_SVC_H_HIGH} => {NPS_Type=Promoter}	0.6228097	0.8185859	1.195179	2026	
[9]	{Tranquility_H=Tranquility_H_HIGH, Check_In_H=Check_In_H_HIGH} => {NPS_Type=Promoter}	0.6108208	0.8080521	1.179800	1987	
[10]	{Guest_Room_H=Guest_Room_H_HIGH, Staff_Cared_H=Staff_Cared_H_HIGH} => {NPS_Type=Promoter}	0.6369505	0.8412505	1.228271	2072	
[11]	{Guest_Room_H=Guest_Room_H_HIGH, Customer_SVC_H=Customer_SVC_H_HIGH} => {NPS_Type=Promoter}	0.6474024	0.8252351	1.204888	2106	
[12]	{Guest_Room_H=Guest_Room_H_HIGH, Check_In_H=Check_In_H_HIGH} => {NPS_Type=Promoter}	0.6344912	0.8174257	1.193486	2064	

V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (2)

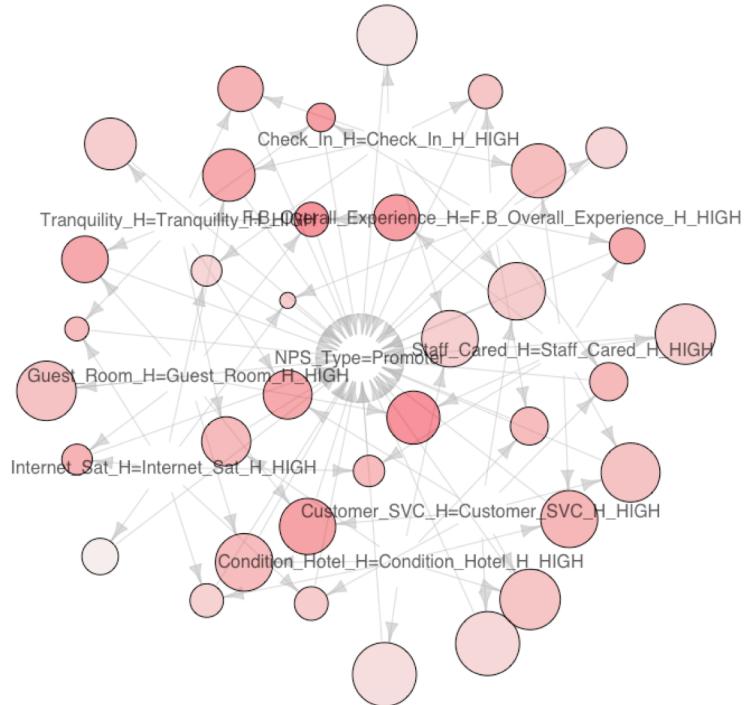
2. Association Rules

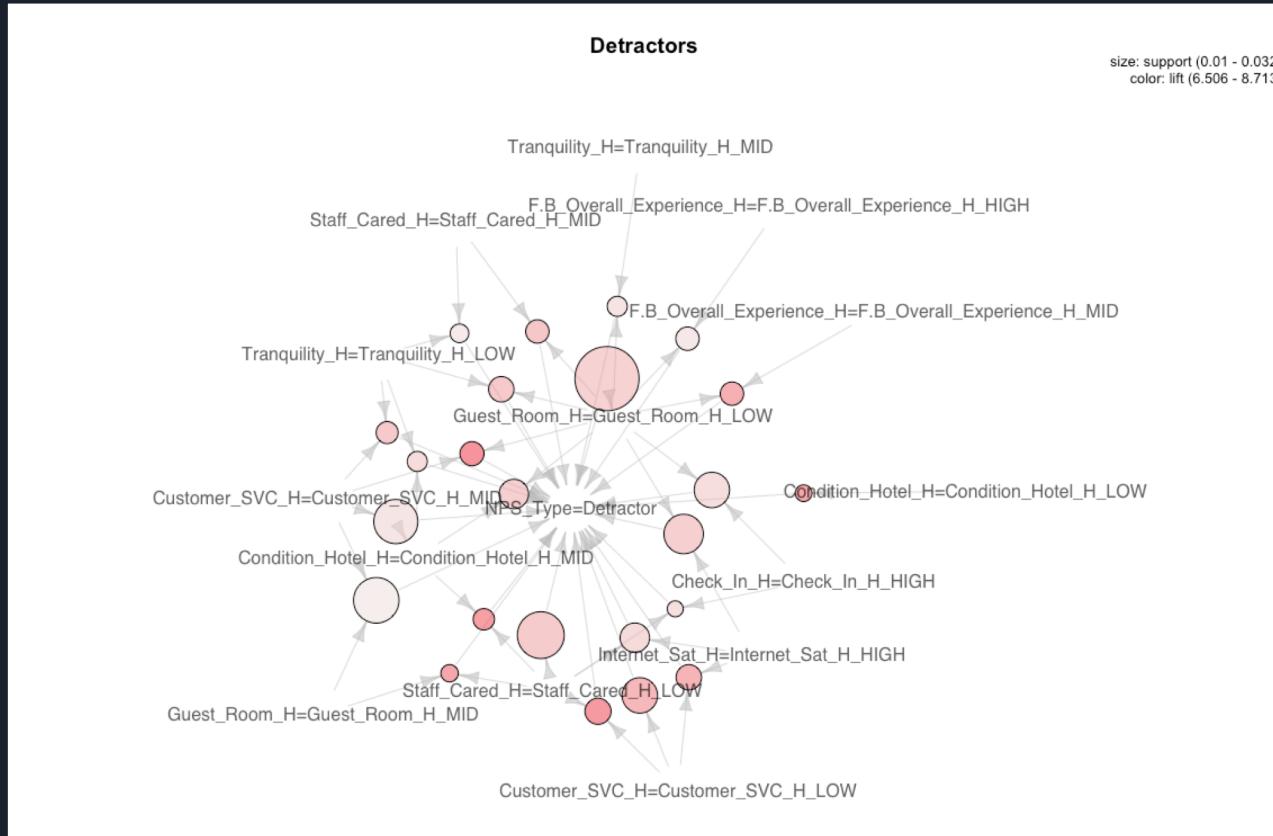
Rules for Detractors:

> inspect(RulesDetractors)	lhs	rhs	support	confidence	lift	count
	[1] {Condition_Hotel_H=Condition_Hotel_H_LOW}	=> {NPS_Type=Detractor}	0.01045189	0.8947368	8.315940	34
	[2] {Customer_SVC_H=Customer_SVC_H_LOW}	=> {NPS_Type=Detractor}	0.01905933	0.8611111	8.003413	62
	[3] {Staff_Cared_H=Staff_Cared_H_LOW}	=> {NPS_Type=Detractor}	0.02428528	0.8229167	7.648423	79
	[4] {Guest_Room_H=Guest_Room_H_LOW}	=> {NPS_Type=Detractor}	0.03227790	0.8015267	7.449618	105
	[5] {Customer_SVC_H=Customer_SVC_H_LOW, Staff_Cared_H=Staff_Cared_H_LOW}	=> {NPS_Type=Detractor}	0.01475561	0.9230769	8.579341	48
	[6] {Customer_SVC_H=Customer_SVC_H_LOW, Internet_Sat_H=Internet_Sat_H_HIGH}	=> {NPS_Type=Detractor}	0.01444820	0.8703704	8.089471	47
	[7] {Condition_Hotel_H=Condition_Hotel_H_MID, Staff_Cared_H=Staff_Cared_H_LOW}	=> {NPS_Type=Detractor}	0.01260375	0.9111111	8.468127	41
	[8] {Guest_Room_H=Guest_Room_H_MID, Staff_Cared_H=Staff_Cared_H_LOW}	=> {NPS_Type=Detractor}	0.01075930	0.8974359	8.341026	35
	[9] {Guest_Room_H=Guest_Room_H_LOW, Tranquility_H=Tranquility_H_LOW}	=> {NPS_Type=Detractor}	0.01444820	0.8245614	7.663709	47
	[10] {Guest_Room_H=Guest_Room_H_LOW, Condition_Hotel_H=Condition_Hotel_H_MID}	=> {NPS_Type=Detractor}	0.01629265	0.8030303	7.463593	53
	[11] {Guest_Room_H=Guest_Room_H_LOW, Customer_SVC_H=Customer_SVC_H_MID}	=> {NPS_Type=Detractor}	0.01383338	0.9375000	8.713393	45
	[12] {Guest_Room_H=Guest_Room_H_LOW, Staff_Cared_H=Staff_Cared_H_MID}	=> {NPS_Type=Detractor}	0.01352598	0.8301887	7.716011	44
	[13] {Guest_Room_H=Guest_Room_H_LOW, F.B_Overall_Experience_H=F.B_Overall_Experience_H_MID}	=> {NPS_Type=Detractor}	0.01352598	0.8800000	8.178971	44

Promoters

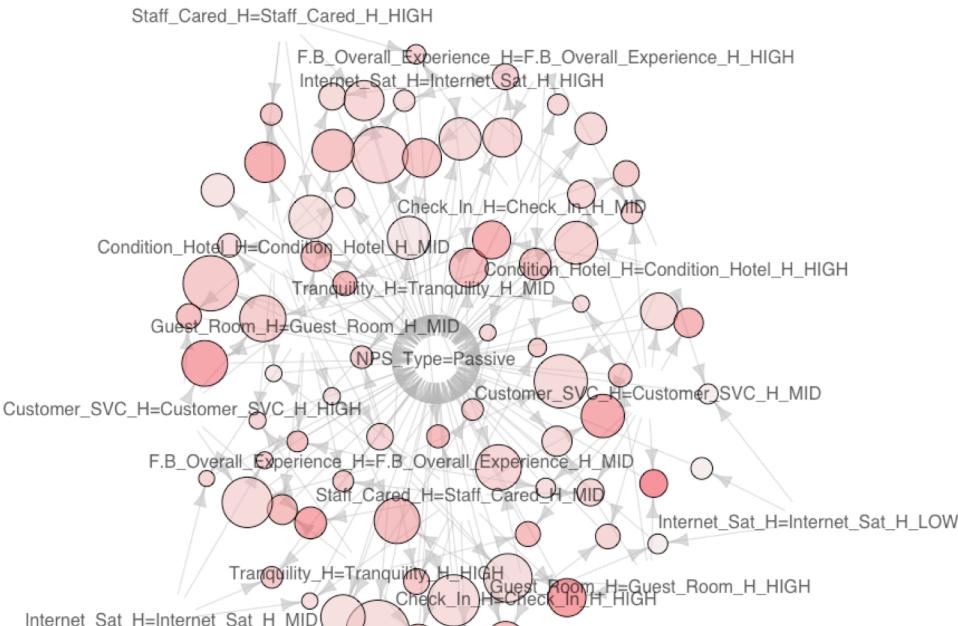
size: support (0.513 - 0.673)
color: lift (1.049 - 1.23)





Passive

size: support (0.01 - 0.064)
color: lift (1.452 - 2.673)



V. DATA MODELING TECHNIQUES & PREDICTIVE ANALYSIS (3)

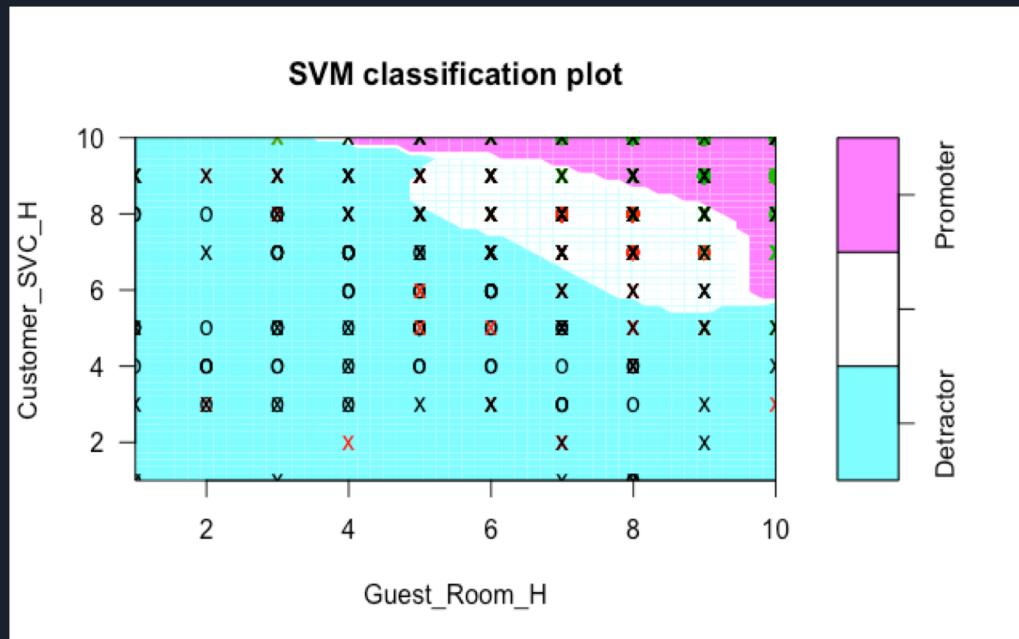
3. Support Vector Machines

View(cor(dataset[,10:18]))

	Overall_Sat_H	Guest_Room_H	Tranquility_H	Condition_Hotel_H	Customer_SVC_H	Staff_Cared_H	Internet_Sat_H	Check_In_H	F.B_Overall_Experience_H
Overall_Sat_H	1.000000	0.7385044	0.6088581	0.6801870	0.7445344	0.6784438	0.2401780	0.5003796	0.4065557
Guest_Room_H	0.7385044	1.000000	0.6476170	0.7282980	0.5577648	0.5129001	0.2172725	0.3907463	0.3398665
Tranquility_H	0.6088581	0.6476170	1.000000	0.5717308	0.5104585	0.4869508	0.1999145	0.3675522	0.3052259
Condition_Hotel_H	0.6801870	0.7282980	0.5717308	1.000000	0.5815353	0.5516152	0.1950891	0.4594733	0.3627324
Customer_SVC_H	0.7445344	0.5577648	0.5104585	0.5815353	1.000000	0.8367528	0.2118969	0.5685682	0.4154054
Staff_Cared_H	0.6784438	0.5129001	0.4869508	0.5516152	0.8367528	1.000000	0.2243039	0.5357339	0.4247022
Internet_Sat_H	0.2401780	0.2172725	0.1999145	0.1950891	0.2118969	0.2243039	1.000000	0.1703789	0.2194686
Check_In_H	0.5003796	0.3907463	0.3675522	0.4594733	0.5685682	0.5357339	0.1703789	1.000000	0.2742025
F.B_Overall_Experience_H	0.4065557	0.3398665	0.3052259	0.3627324	0.4154054	0.4247022	0.2194686	0.2742025	1.000000

SVM Classification

- SVM separated the plots into three parts: Detractor, Passive, and Promoter.
- we can see that SVM worked almost successfully.
- Most black points are in the Detractor area; most red points are in the Passive area; most green points are located in the Promoter area.





SVM Analysis and Recommendation

Analysis:

We performed SVM and got the same results as per our linear modeling analysis. This proves that our linear modeling is correct. According to SVM, Customer_SVC_H is more important than Guest_Room_H. The detractor area includes all the part, when the Guest_Room_H scale is from 0-10. On the contrary, when Customer_SVC_H changes, it greatly affects the NPS_Type.

Recommendation:

- Based on SVM Classification analysis Hyatt hotel should pay close attention on Customer service.
- Provide more customer service training for management and staff.
- Cultivate and maintain a strong customer-oriented experience by way of providing exceptional personal (positive employee-to-customer interaction) service.



VI. DATA GENERALIZATION

- Reran our 3 predictive modeling algorithms on more data (the U.S.) to compare the results and get higher accuracies
- We found similar results!



VII. DATA VALIDATION

- Checked one another's work and code on a regular basis
- Used different tools to make sure our calculations were always right
- Our `ggplot2` visualizations were always verified through Tableau Desktop, due to the ease of use of the program (i.e. lots of drag and dropping, and quick preliminary results.) We made compared the results and made sure our Tableau and R visualizations were similar.

VIII. RESULTS

Primary (for quality purposes)

- Focus on all of the variables relating to customer service but put the most emphasis specifically on the top two variables: Customer_SVC_H and Guest_Room_H.
- Put less emphasis on Tranquility_H and F.B_Overall_Experience_H as they have less of an effect, but do not disregard them altogether as they do have a little bit of influence.
- Do not focus on variables such as Internet_Sat_H and Check_in_H, they don't have much if any influence on customer satisfaction.
- Based on SVM Classification analysis, Hyatt hotels should pay close attention to all things Customer service.
- Provide more customer service training for management and staff.
- Incentivize giving exceptional service by rewarding employees for good reviews.
- Cultivate, but also work to maintain a strong customer-oriented experience by way of providing exceptional personal (positive employee-to-customer interaction) service at all times.

Secondary (for marketing purposes)

- NYC is an international tourist hub, as such, it is disappointing that the majority of the customers are U.S.-based travellers. The Hyatt Hotel Corporation should try to attract more international travellers.
- Hyatt Hotel Corporation should consider a wider variety of demographics:
 - Female travelers
 - Leisure travelers
 - Younger and older travelers besides the core demographic (40-60)

Thank you

