UNIVERSITY OF TEXAS AT ARLINGTON
INSY-5339-003-PRIN OF BUSINESS DATA MINING
SPRING 2022

FINAL PROJECT REPORT

Due Date: 5/8/2022

Submitted by: Team 9

Team Members

Ruchi Shukla: 1001977095

Pranati Chauhan: 1002032137

Jessica Mendem: 1002036921

INDEX

Executive Summary

Company Goal:

Boost business for hotels by predicting cancellations.

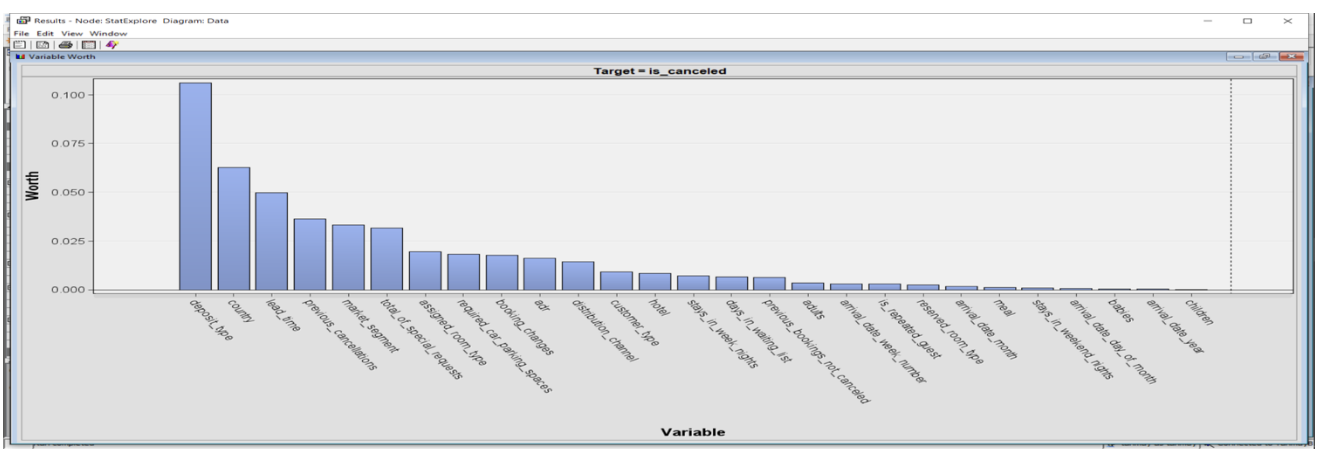
Problem Statement:

In the hospitality industry, there is a vast amount of data which is untouched that can revolutionize revenue management. In the industry, optimizing the operations of the hotel industry to ensure the right room to the right customer at the right price is what would hotels stay ahead of the competition and help in efficient use of resources.

Last-minute cancellations and no-show lead to heavy loss of resources. What if cancellations could be predicted way before they are made?

Proposed Solution:

If potential cancellers are identified, they may be persuaded to check into a hotel by attractive, customized incentives. Analyzing the behavior of previous hotel guests might help management provide targeted incentives, allowing them to keep as many reservations as feasible.



As seen in the bar graph, it can be concluded that deposit, country, and lead time are the top 3 factors which lead to cancellations.

A proper solution investigating these factors can help to minimize cancellations.

Approach:

To implement the concept, data was analyzed from three years (2015-2017) of customer records at city and resort hotels. To construct the best accurate prediction model for identifying probable cancellers and studying prior customer behavior, a variety of data mining algorithms were applied.

Results:

To detect prospective cancellers, a predictive model with a high degree of accuracy was successfully created. Various parameters were used to assess consumer behavior and create client categories. Incentives were given to customers who were likely to cancel reservations.

Managerial Insights:

Managers might utilize the prediction model to target possible cancellers, and group of customers analysis could be used to give individualized rewards to the identified potential cancellers. The benefits must be based on the resources and capital available.

Hotel Booking Cancellations

Background:

We are a data analytics team for a hotel chain, and our job is to provide management with insights that will help them enhance business at their city and resort properties.

• Numerous individuals who reserve hotel reservations end up checking in, but there are also many people who cancel their reservations.

• It is wonderful for business if people show up; but it is a loss for the franchise if they cancel their reservations and do not show up.

Solution:

Management has advised us that if we can identify possible cancellers and provide them with individualized incentives, they may be persuaded to visit a hotel, minimizing the chance of a reservation cancellation.

We'll discover and investigate folks who have a high risk of canceling in order to provide them with targeted incentives.

What Should You Do

- If we can identify potential cancellers, we might be able to convince them and provide specific incentives, minimizing the risk of a reservation cancellation.
- We'll discover and investigate folks who have a high risk of canceling to provide them with targeted incentives.

Description of Data

Resources on hand

Data was collected from Kaggle.

https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand

- This dataset contains client records of 2 hotels- city hotel and a resort hotel (2015-2017).
- This dataset comprises booking information including dates of booking, duration of stay, number of adults, children, and/or newborns, meal preferences, room type, etc.

List of Variables

Variable	**Description**	**Type**
Is_canceled	If booking was cancelled	Categorical
lead_time	difference between booking date & arrival date	Integer
arrival_date_year	arrival date of the year	Integer
arrival_date_month	arrival month - Jan to Dec	Categorical
arrival_date_week_number	week of arrival	Integer
arrival_date_day_of_month	date and day of arrival	Integer
stays_in_weekend_nights	number of nights stayed on weekends	Integer
stays_in_week_nights	number of nights stayed in a week	Integer
Adults	Number of adults	Integer
Children	Number of children	Integer
Babies	Number of babies	Integer

Meal	type of meal booked	Categorical
Country	country of origin of the guest	Categorical
market_segment	source of booking - self or through travel agent or other sources	Categorical
distribution_channel	source of booking - self or through travel agent or other sources	Categorical
is_repeated_guest	If the guest	Categorical
previous_cancellations	bookings cancelled by the customer at the same hotel before the current booking	Integer
previous_bookings_not_canceled	number of bookings not cancelled by guest, prior to current booking	Integer
reserved_room_type	which type of room was booked by the customer	Categorical
assigned_room_type	which room was assigned	Categorical
booking_changes	count of changes made to the reservation ahead of check-in	Integer
deposit_type	Category addressing whether a deposit was made during reservation or not	Categorical
Agent	travel agency identifier/code	Categorical
Company	company identifier/code	Categorical
days_in_waiting_list	measurement of time when booking was on waitlist	Integer
customer_type	type of customer	Categorical

Adr	Average daily rates at the time of booking	Numeric
required_car_parking_spaces	number of parking spaces requested in the booking	Integer
total_of_special_requests	count of special requests associated with the reservation	Integer
reservation_status	status of reservation at the end of stay	Categorical
reservation_status_date	date of reservation_status	Date

Summary of Variables

Total number of records	118,221
Binary variable	2
Ordinal variable	4
Nominal	9
Interval	13
Outcome/Target variable	Is_canceled
Percentage of the records that belong to each class.	For is_canceled = 0(Not Canceled):74283 Observations = 63% For is_canceled = 1(Canceled): 43938 Observations = 37%

Data Visualization

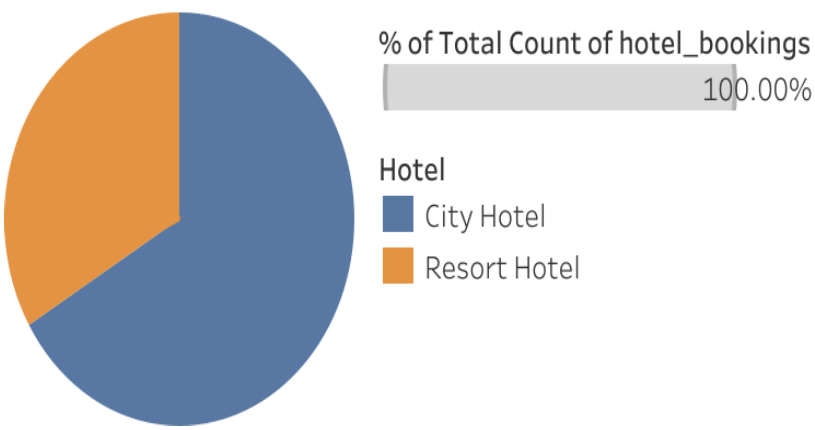
For data exploration and finding trends and patterns, data visualization is a necessary tool. It helps us understand relationships between variables.

Through these visualizations, univariate analysis and various patterns and trends helped in understanding and brainstorming why cancellations happen and how changes can be made by analyzing the trends.

Visualization Methods:

- Bar Chart
- Pie Chart
- Heat Map

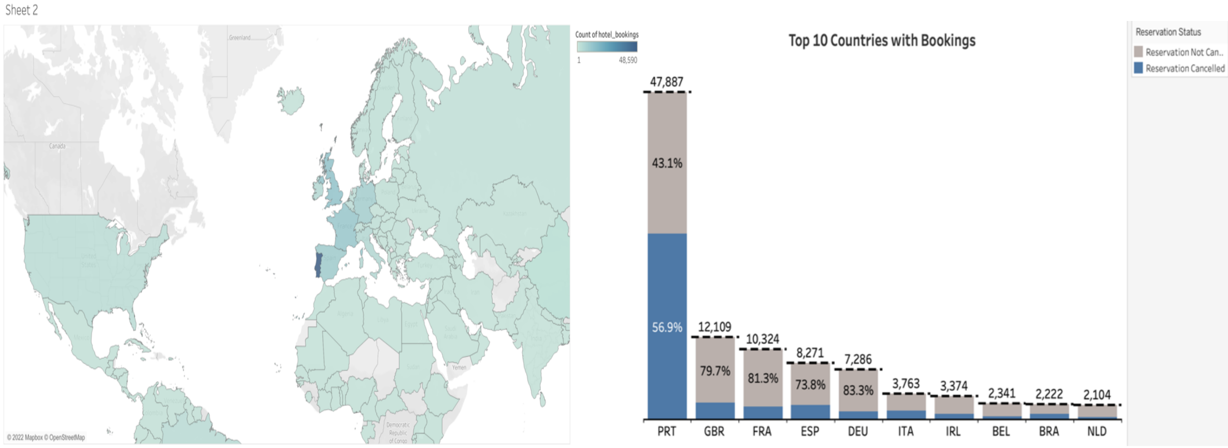
Univariate Analysis



The univariate analysis here shows the following:

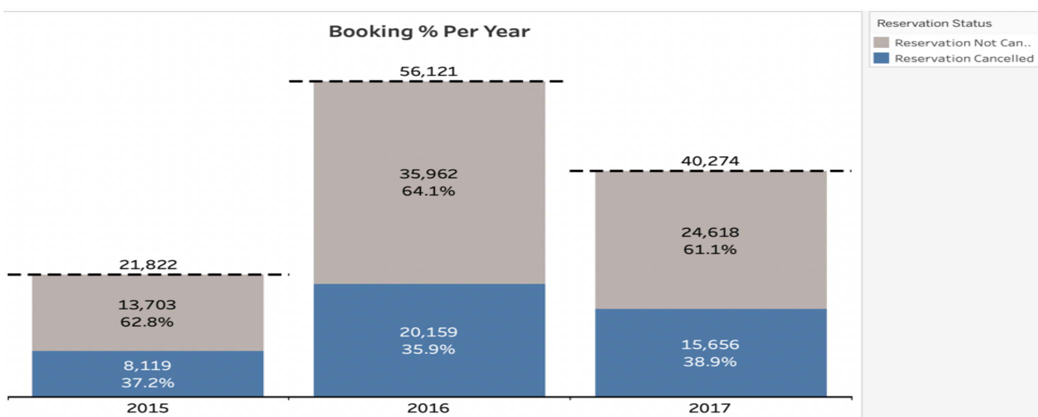
- City Hotel had more reservations and cancellations than Resort Hotels.
- Around 49% of the bookings were made by Transient customer type and had a 69% rate of cancellations in city hotels.
- Resort Hotels had most bookings and cancellations by Transient Customers as well.

Maps

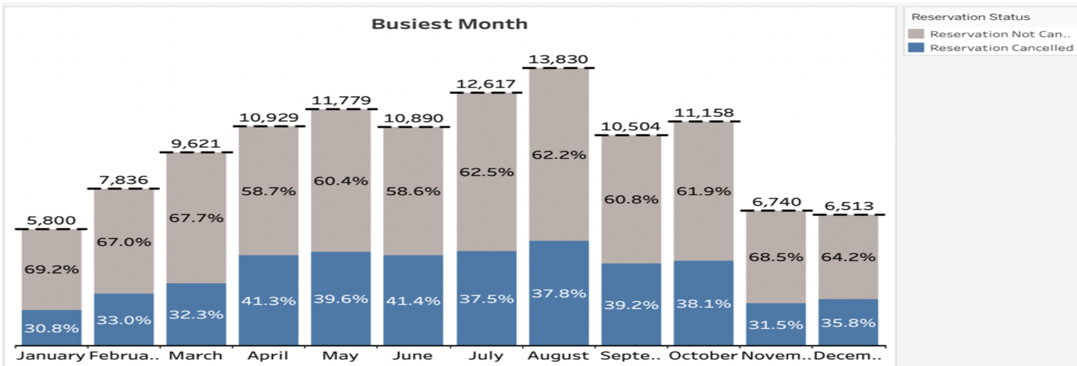


- Portugal has the most bookings and cancellations.
- Among 47.887% of bookings, 56.9% were cancelled, which is an alarming percentage.

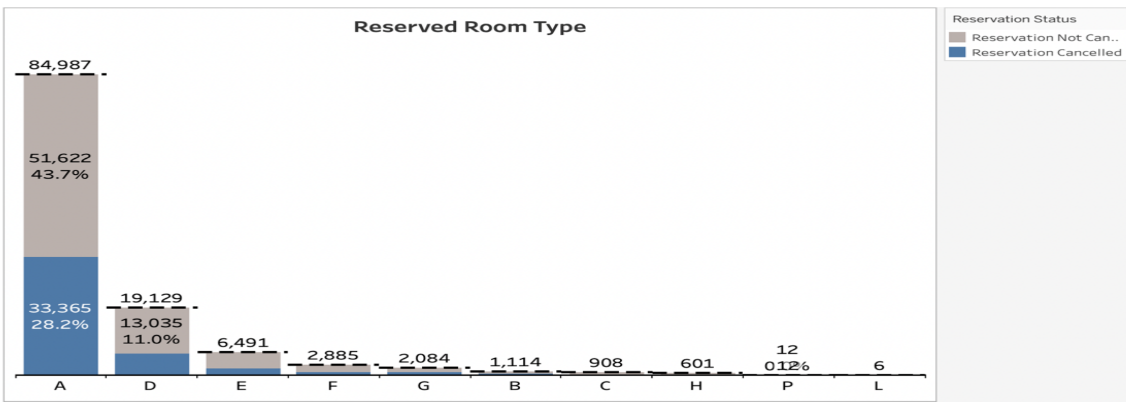
Bar Graph



- **Booking percentage per year:** In 2015 bookings are 62.8% and Cancellation 37.2%. The highest bookings were in 2016 bookings are 64.1% and cancellations are 35.9%. In 2017 bookings were up to 61.1% and 38.9% Cancellations.

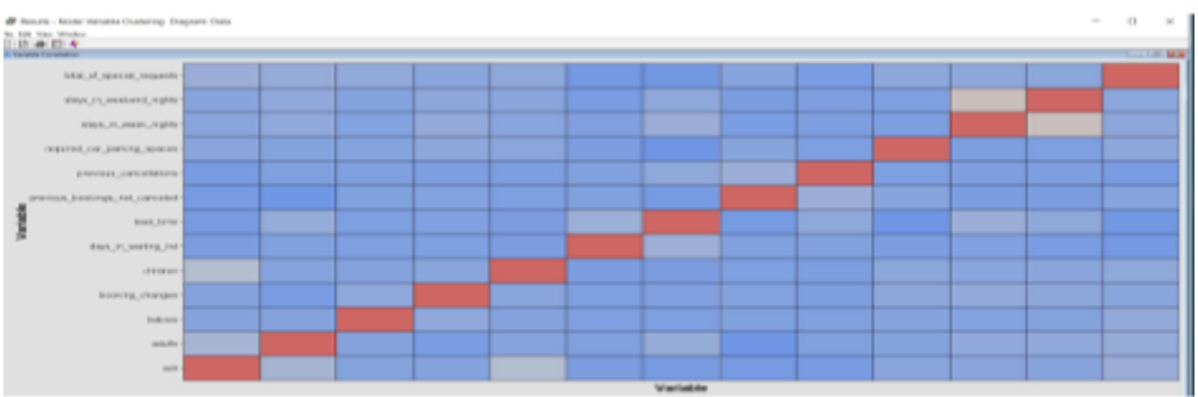


- **Busiest Month:** In overall months we have busiest month is in **August** and it also has the highest cancellation rate.



- Customers who booked **Room A** cancelled more than any other room.

Heat Map



- It was found that the variables were weakly correlated with each other.

Data Mining Models

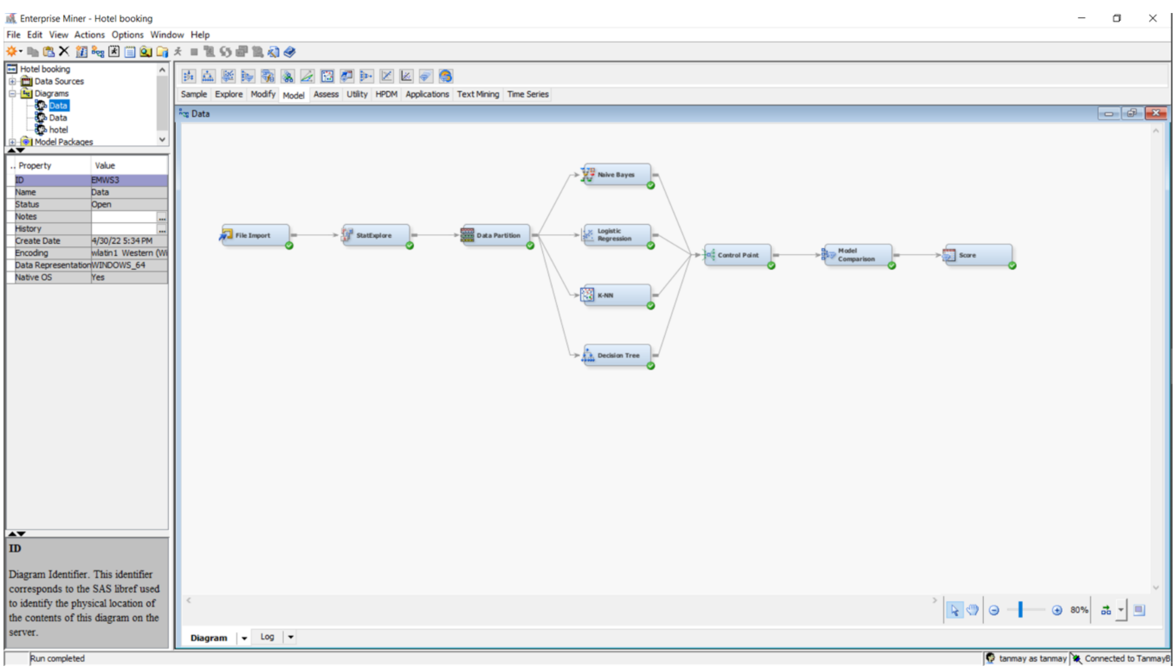
The data was utilized to create a model that predicted how likely a potential customer would cancel their reservation.

For developing the prediction model:

- Naïve Bayes Classification
- K- Nearest Neighbor Algorithm
- Classification Tree
- Logistic Regression

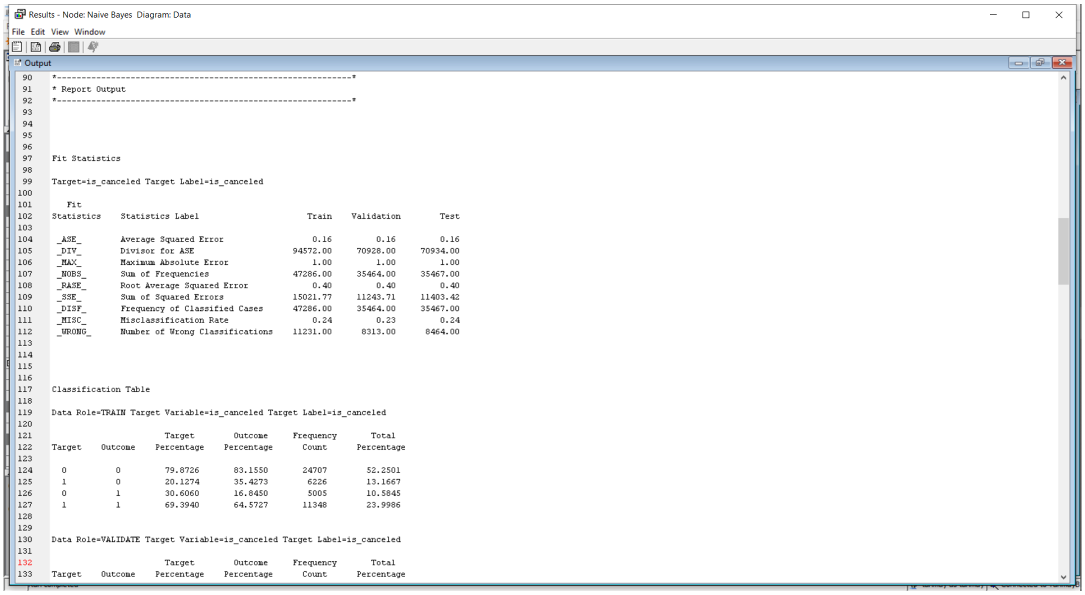
- After partitioning the dataset into a 60:20:20 ratio (Training: Validation: Test), all the models were run.
- **Naive Bayes and K-Nearest Neighbor** algorithms were used, because our dataset was vast, and these algorithms perform better on large datasets.
- In case the above-mentioned models' performance was unsatisfactory, **classification tree and logistic regression** were used as backups.
- We also performed different regression models- stepwise, forward, and backward, and the misclassification rate for forward regression was minimum.

Process Flow Diagram



Model 1: Naïve Bayes Classification:

Results



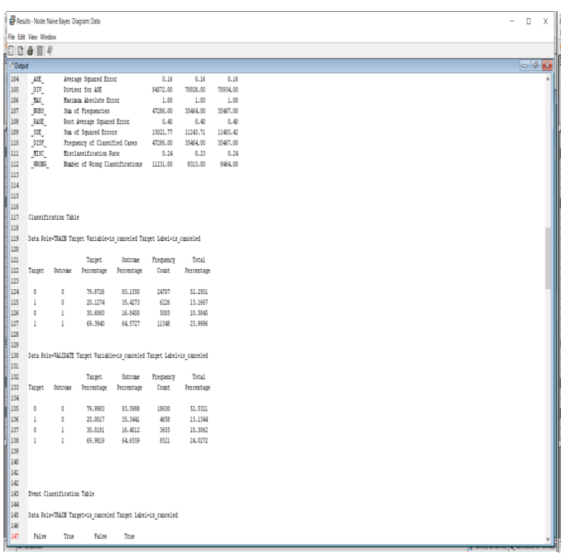
Results - Node: Naive Bayes Diagram: Data

File Edit View Window

Output

```
*------------------------------------------------------------*
* Report Output
*------------------------------------------------------------*



Fit Statistics

Target=is_canceled Target Label=is_canceled

   Fit
Statistics    Statistics Label                    Train    Validation       Test

 _ASE_        Average Squared Error                0.16          0.16       0.16
 _DIV_        Divisor for ASE                  94572.00      70928.00   70934.00
 _MAX_        Maximum Absolute Error               1.00          1.00       1.00
 _NOBS_       Sum of Frequencies               47286.00      35464.00   35467.00
 _RASE_       Root Average Squared Error           0.40          0.40       0.40
 _SSE_        Sum of Squared Errors            15021.77      11243.71   11403.42
 _DISF_       Frequency of Classified Cases    47286.00      35464.00   35467.00
 _MISC_       Misclassification Rate               0.24          0.23       0.24
 _WRONG_      Number of Wrong Classifications  11231.00       8313.00    8464.00



Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome     Frequency      Total
Target    Outcome    Percentage   Percentage     Count      Percentage

  0          0        79.8726      83.1550       24707       52.2501
  1          0        20.1274      35.4273        6226       13.1667
  0          1        30.6060      16.8450        5005       10.5845
  1          1        69.3940      64.5727       11348       23.9986


Data Role=VALIDATE Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome     Frequency      Total
Target    Outcome    Percentage   Percentage     Count      Percentage
```

Statistics and Bayesian Network

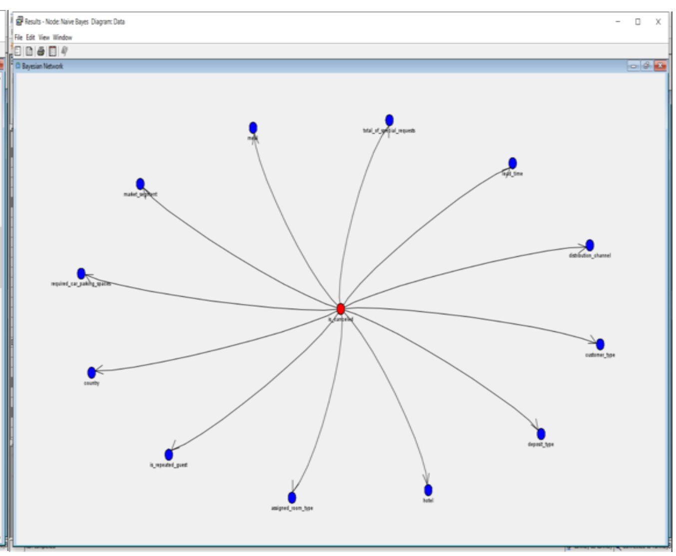


Results - Node: Naive Bayes Diagram: Data

File Edit View Window

Output

```
 _ASE_        Average Squared Error                0.16          0.16       0.16
 _DIV_        Divisor for ASE                  94572.00      70928.00   70934.00
 _MAX_        Maximum Absolute Error               1.00          1.00       1.00
 _NOBS_       Sum of Frequencies               47286.00      35464.00   35467.00
 _RASE_       Root Average Squared Error           0.40          0.40       0.40
 _SSE_        Sum of Squared Errors            15021.77      11243.71   11403.42
 _DISF_       Frequency of Classified Cases    47286.00      35464.00   35467.00
 _MISC_       Misclassification Rate               0.24          0.23       0.24
 _WRONG_      Number of Wrong Classifications  11231.00       8313.00    8464.00



Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome     Frequency      Total
Target    Outcome    Percentage   Percentage     Count      Percentage

  0          0        79.8726      83.1550       24707       52.2501
  1          0        20.1274      35.4273        6226       13.1667
  0          1        30.6060      16.8450        5005       10.5845
  1          1        69.3940      64.5727       11348       23.9986


Data Role=VALIDATE Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome     Frequency      Total
Target    Outcome    Percentage   Percentage     Count      Percentage

  0          0        79.9983      83.5988       18630       52.5321
  1          0        20.0017      35.3441        4658       13.1344
  0          1        30.0181      16.4012        3655       10.3062
  1          1        69.9819      64.6559        8521       24.0272



Event Classification Table

Data Role=TRAIN Target=is_canceled Target Label=is_canceled

  False      True       False      True
```



Observations

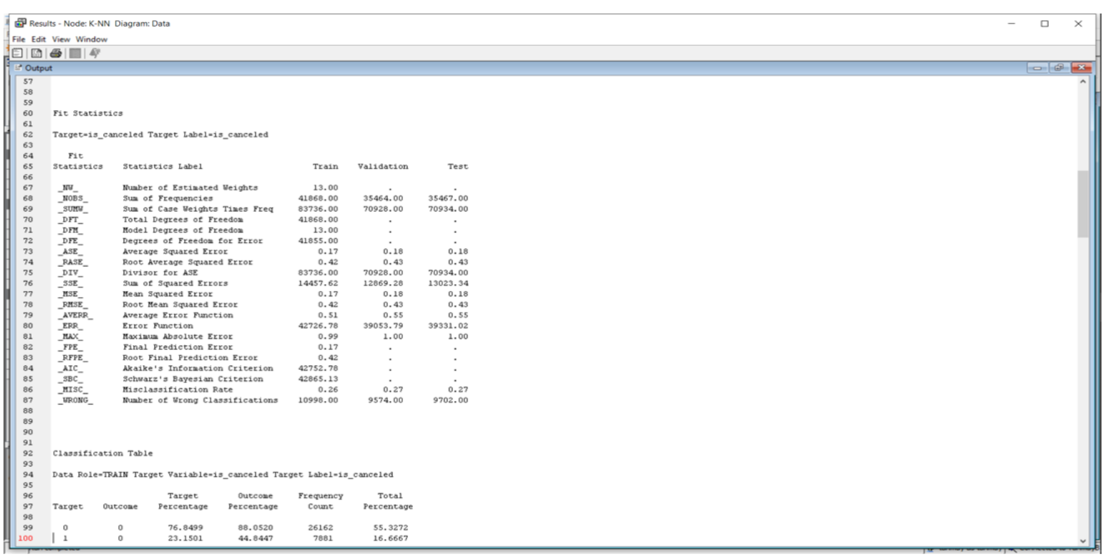
- Average Squared Error for validation dataset **= 17%**
- Misclassification Rate for validation dataset **= 24%**

Important Variables in Bayesian Network:

- lead_time
- market_segment
- assigned_room_type
- customer_type
- required_car_parking_spaces
- country
- hotel
- is_repeated_guest
- meal
- reserved_room_type
- deposit_type
- distribution_channel
- total_of_special_requests

Model 2: KNN Algorithm - Default settings were used.

Results



Results - Node: K-NN Diagram: Data

File Edit View Window

Output

Fit Statistics

Target=is_canceled Target Label=is_canceled

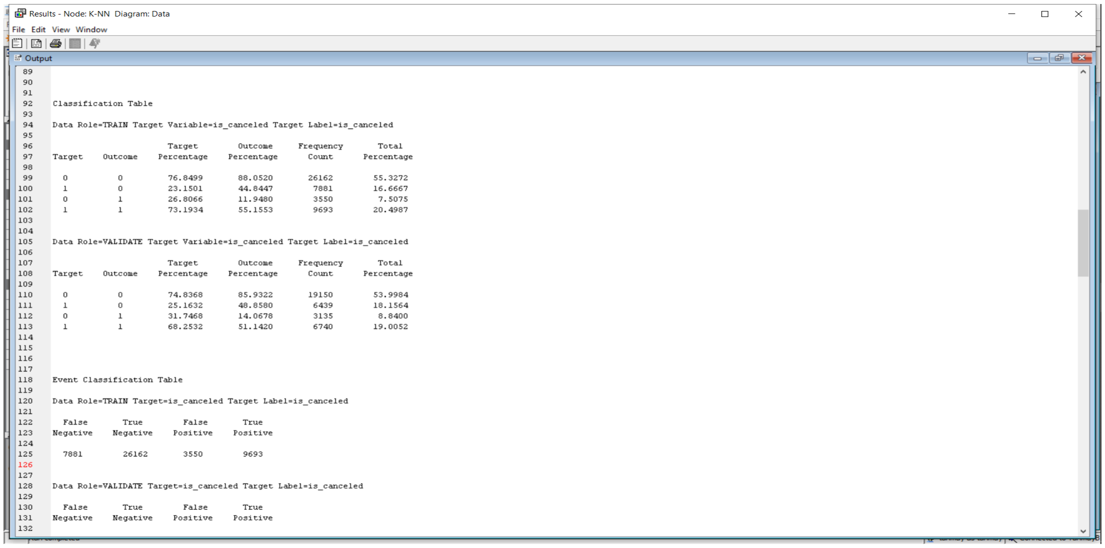
Fit Statistics	Statistics Label	Train	Validation	Test
_NW_	Number of Estimated Weights	13.00	.	.
NOBS	Sum of Frequencies	41868.00	35464.00	35467.00
SUMW	Sum of Case Weights Times Freq	83736.00	70928.00	70934.00
DFT	Total Degrees of Freedom	41868.00	.	.
DFM	Model Degrees of Freedom	13.00	.	.
DFE	Degrees of Freedom for Error	41855.00	.	.
ASE	Average Squared Error	0.17	0.18	0.18
RASE	Root Average Squared Error	0.42	0.43	0.43
DIV	Divisor for ASE	83736.00	70928.00	70934.00
SSE	Sum of Squared Errors	14457.62	12869.28	13023.34
MSE	Mean Squared Error	0.17	0.18	0.18
RMSE	Root Mean Squared Error	0.42	0.43	0.43
AVERR	Average Error Function	0.51	0.55	0.55
ERR	Error Function	42726.78	39053.79	39331.02
MAX	Maximum Absolute Error	0.99	1.00	1.00
FPE	Final Prediction Error	0.17	.	.
RFPE	Root Final Prediction Error	0.42	.	.
AIC	Akaike's Information Criterion	42752.78	.	.
SBC	Schwarz's Bayesian Criterion	42865.13	.	.
MISC	Misclassification Rate	0.26	0.27	0.27
WRONG	Number of Wrong Classifications	10998.00	9574.00	9702.00

Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	76.8499	88.0520	26162	55.3272
1	0	23.1501	44.8447	7881	16.6667

Confusion Matrix



Results - Node: K-NN Diagram: Data

File Edit View Window

Output

Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	76.8499	88.0520	26162	55.3272
1	0	23.1501	44.8447	7881	16.6667
0	1	26.8066	11.9480	3550	7.5075
1	1	73.1934	55.1553	9693	20.4987

Data Role=VALIDATE Target Variable=is_canceled Target Label=is_canceled

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	74.8368	85.9322	19150	53.9984
1	0	25.1632	48.8580	6439	18.1564
0	1	31.7468	14.0678	3135	8.8400
1	1	68.2532	51.1420	6740	19.0052

Event Classification Table

Data Role=TRAIN Target=is_canceled Target Label=is_canceled

False Negative	True Negative	False Positive	True Positive
7881	26162	3550	9693

Data Role=VALIDATE Target=is_canceled Target Label=is_canceled

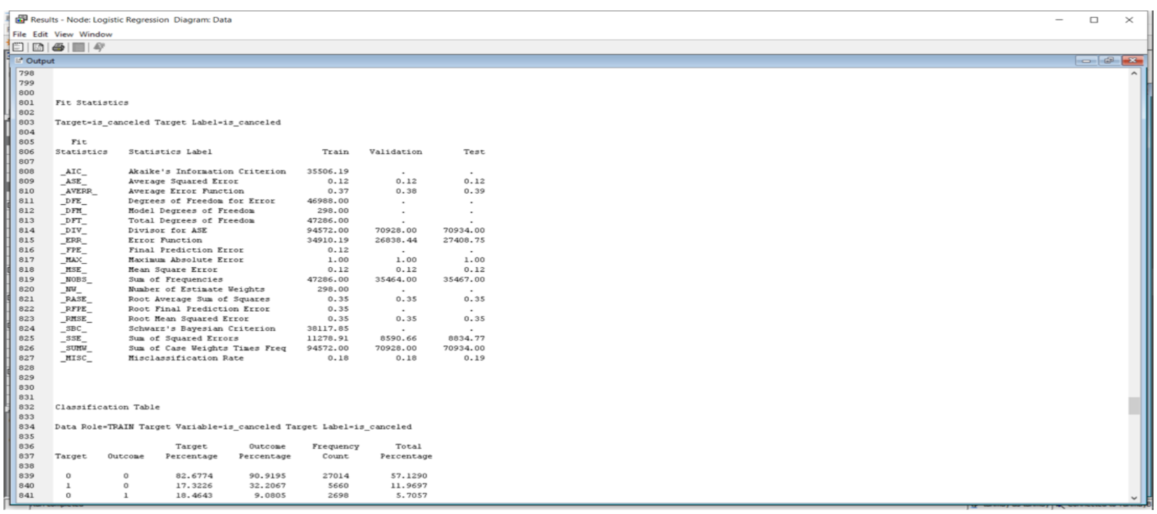
False Negative	True Negative	False Positive	True Positive

Observations

- Average Squared Error for validation dataset **= 18%**
- Misclassification Rate for validation dataset **= 26%**

Model 3: Classification Tree- Default settings were used.

Results



Results - Node: Logistic Regression Diagram: Data

File Edit View Window

Output

Fit Statistics

Target=is_canceled Target Label=is_canceled

Fit Statistics	Statistics Label	Train	Validation	Test
AIC	Akaike's Information Criterion	35506.19	.	.
ASE	Average Squared Error	0.12	0.12	0.12
AVERR	Average Error Function	0.37	0.38	0.39
DFE	Degrees of Freedom for Error	46988.00	.	.
DFM	Model Degrees of Freedom	298.00	.	.
DFT	Total Degrees of Freedom	47286.00	.	.
DIV	Divisor for ASE	94572.00	70928.00	70934.00
ERR	Error Function	34910.19	26838.44	27408.75
FPE	Final Prediction Error	0.12	.	.
MAX	Maximum Absolute Error	1.00	1.00	1.00
MSE	Mean Square Error	0.12	0.12	0.12
NOBS	Sum of Frequencies	47286.00	35464.00	35467.00
_NW_	Number of Estimate Weights	298.00	.	.
RASE	Root Average Sum of Squares	0.35	0.35	0.35
RFPE	Root Final Prediction Error	0.35	.	.
RMSE	Root Mean Squared Error	0.35	0.35	0.35
SBC	Schwarz's Bayesian Criterion	38117.85	.	.
SSE	Sum of Squared Errors	11278.91	8590.66	8834.77
SUMW	Sum of Case Weights Times Freq	94572.00	70928.00	70934.00
MISC	Misclassification Rate	0.18	0.18	0.19

Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	82.6774	90.9195	27014	57.1290
1	0	17.3226	32.2067	5660	11.9697
0	1	18.4643	9.0805	2698	5.7057

Observations:

- Average Squared Error for validation dataset **= 13%**
- Misclassification Rate for validation dataset **= 20%**

List of important variables

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
deposit_type		1	1.0000	1.0000	1.0000
country		2	0.4782	0.4715	0.9861
market_segment		1	0.4289	0.4214	0.9824
lead_time		3	0.3975	0.4041	1.0166
total_of_special_requests		1	0.3444	0.3642	1.0577
previous_cancellations		3	0.2811	0.3006	1.0695
required_car_parking_spaces		1	0.2574	0.2625	1.0200
booking_changes		1	0.1868	0.2015	1.0787
arrival_date_year		1	0.1484	0.1412	0.9516
previous_bookings_not_canceled		2	0.0786	0.0858	1.0925
distribution_channel		1	0.0710	0.0608	0.8556
customer_type		1	0.0513	0.0537	1.0461
hotel		1	0.0337	0.0292	0.8658

Model 4: Logistic Regression – Default settings were used.

Results

Fit Statistics

Target=is_canceled Target Label=' '

Fit Statistics	Statistics Label	Train	Validation	Test
AIC	Akaike's Information Criterion	53505.06	.	.
ASE	Average Squared Error	0.12	0.12	0.12
AVERR	Average Error Function	0.37	0.39	0.38
DFE	Degrees of Freedom for Error	70696.00	.	.
DFM	Model Degrees of Freedom	235.00	.	.
DFT	Total Degrees of Freedom	70931.00	.	.
DIV	Divisor for ASE	141862.00	47290.00	47290.00
ERR	Error Function	53035.06	18289.20	18145.51
FPE	Final Prediction Error	0.12	.	.
MAX	Maximum Absolute Error	1.00	1.00	1.00
MSE	Mean Square Error	0.12	0.12	0.12
NOBS	Sum of Frequencies	70931.00	23645.00	23645.00
_NW_	Number of Estimate Weights	235.00	.	.
RASE	Root Average Sum of Squares	0.35	0.35	0.35
RFPE	Root Final Prediction Error	0.35	.	.
RMSE	Root Mean Squared Error	0.35	0.35	0.35
SBC	Schwarz's Bayesian Criterion	55659.89	.	.
SSE	Sum of Squared Errors	17216.67	5819.42	5870.64
SUMW	Sum of Case Weights Times Freq	141862.00	47290.00	47290.00
MISC	Misclassification Rate	0.18	0.18	0.18

Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=' '

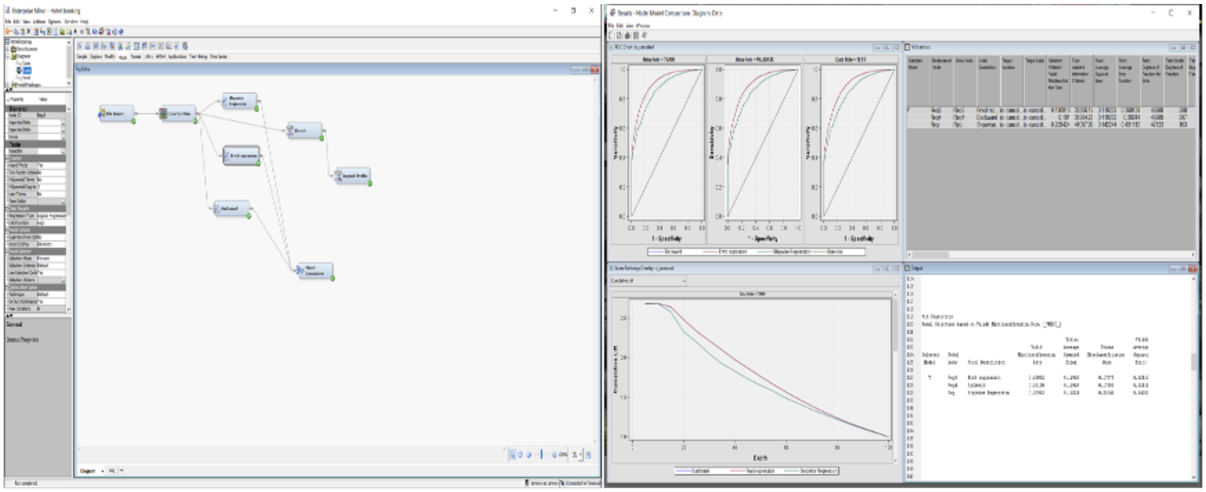
Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	82.1738	90.7694	40455	57.0343

.

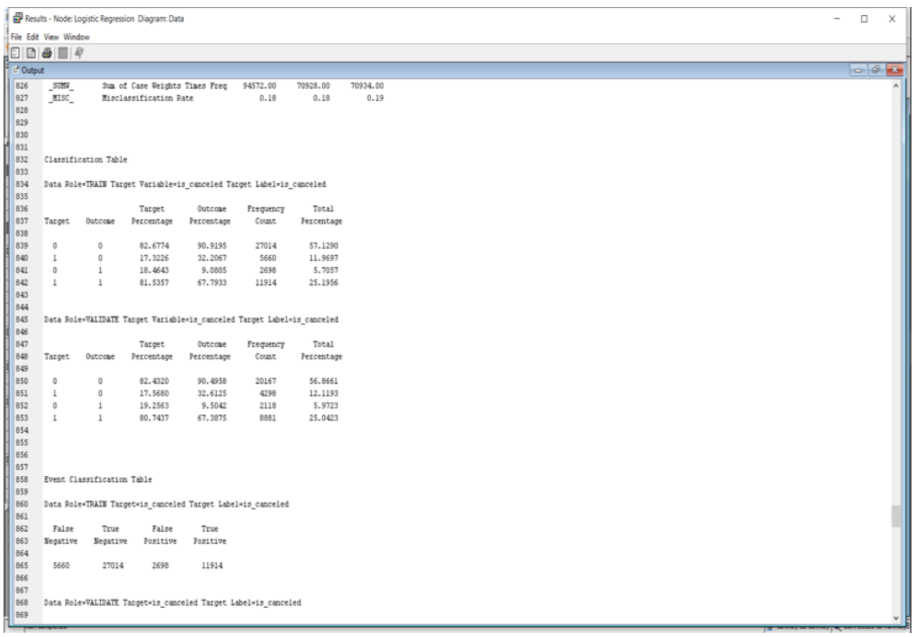
Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
IMP_children	1	35.4431	<.0001
adr	1	346.1667	<.0001
adults	1	49.1659	<.0001
arrival_date_day_of_month	1	57.7395	<.0001
arrival_date_month	11	293.6240	<.0001
arrival_date_week_number	1	60.3100	<.0001
arrival_date_year	0	0.0000	.
assigned_room_type	11	922.4156	<.0001
babies	1	0.5301	0.4666
booking_changes	1	262.8020	<.0001
country	168	5314.6726	<.0001
customer_type	3	388.2299	<.0001
days_in_waiting_list	1	13.1290	0.0003
deposit_type	2	989.6810	<.0001
distribution_channel	3	717.1956	<.0001
hotel	1	5.2996	0.0213
is_repeated_guest	1	104.8284	<.0001
lead_time	1	1730.1400	<.0001
market_segment	7	1665.8151	<.0001
meal	3	92.7265	<.0001
previous_bookings_not_canceled	1	164.7710	<.0001
previous_cancellations	1	706.1083	<.0001
required_car_parking_spaces	1	17.7431	<.0001
reserved_room_type	8	631.7160	<.0001
stays_in_week_nights	1	90.9505	<.0001
stays_in_weekend_nights	1	64.6551	<.0001
total_of_special_requests	1	2205.5688	<.0001

Comparison between different Regression



Confusion Matrix



Results - Node: Logistic Regression Diagram: Data

File Edit View Window

Output

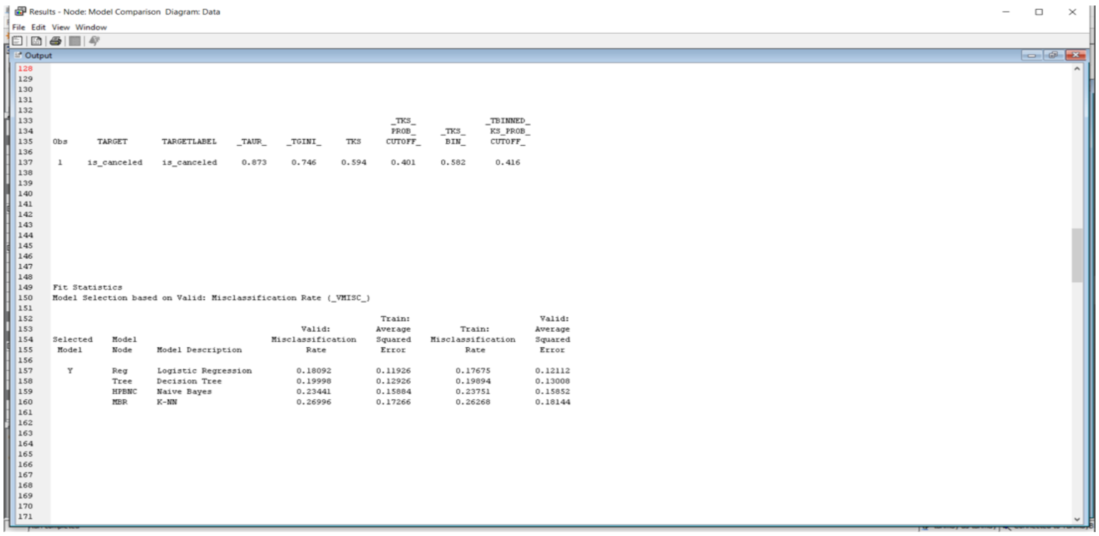
```
 _SUMW_      Sum of Case Weights Times Freq    94572.00     70928.00     70934.00
 _MISC_      Misclassification Rate                0.18         0.18         0.19




Classification Table

Data Role=TRAIN Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome    Frequency     Total
Target    Outcome    Percentage   Percentage    Count     Percentage

  0          0        82.6774      90.9195      27014      57.1290
  1          0        17.3226      32.2067       5660      11.9697
  0          1        18.4643       9.0805       2698       5.7057
  1          1        81.5357      67.7933      11914      25.1956


Data Role=VALIDATE Target Variable=is_canceled Target Label=is_canceled

                       Target      Outcome    Frequency     Total
Target    Outcome    Percentage   Percentage    Count     Percentage

  0          0        82.4320      90.4958      20167      56.8661
  1          0        17.5680      32.6125       4298      12.1193
  0          1        19.2563       9.5042       2118       5.9723
  1          1        80.7437      67.3875       8881      25.0423




Event Classification Table

Data Role=TRAIN Target=is_canceled Target Label=is_canceled

  False      True       False      True
Negative   Negative   Positive   Positive

  5660      27014       2698      11914


Data Role=VALIDATE Target=is_canceled Target Label=is_canceled
```

Observations

- Average Squared Error for validation dataset **= 12%**
- Misclassification Rate for validation dataset **= 18%**

Model Comparison

- K-NN model was predicting incorrectly.
- Logistic regression and classification tree were chosen since they performed much better.
- Misclassification rate for the logistic regression model was **18%.**


Results - Node: Model Comparison Diagram: Data

File Edit View Window

Output

Obs	TARGET	TARGETLABEL	_TAUR_	_TGINI_	TKS	_TKS_ PROB_ CUTOFF_	_TKS_ BIN_	_TBINNED_ KS_PROB_ CUTOFF_
1	is_canceled	is_canceled	0.873	0.746	0.594	0.401	0.582	0.416

Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

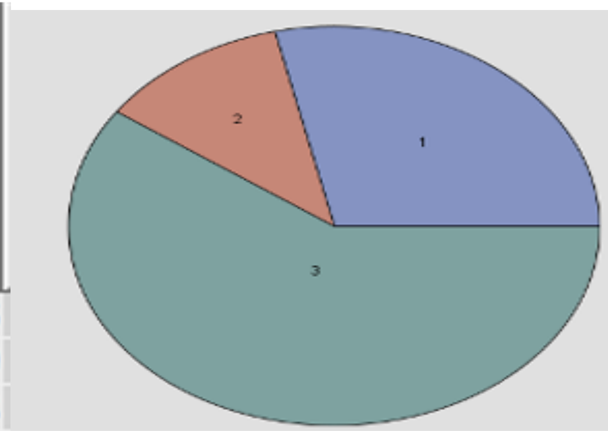
Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Reg	Logistic Regression	0.18092	0.11926	0.17675	0.12112
	Tree	Decision Tree	0.19998	0.12926	0.19894	0.13008
	HPBNC	Naive Bayes	0.23441	0.15884	0.23751	0.15852
	MBR	K-NN	0.26996	0.17266	0.26268	0.18144

Clustering - 3 significant clusters were formed: Cluster 1, 2, 3

Cluster 2: Customers likely to cancel (1)

Cluster 3: Customers not likely to cancel (0)

is_canceled=0	is_canceled=1
0.540964	0.459036
0.351651	0.648349
0.725227	0.274773



Segment 2: 65% consists of people who canceled reservations.

Statistics for variables:

- Lead time – 329
- Market segment – Groups = 48%, Offline TA/TO = 31%
- Deposit Type – No Deposit = 59%
- Average Daily Rate = $84 .00
- Late check ins
- Highest number of previous cancellations
- Travelers were mostly from **Portugal** (59%) and **Germany** (10%)
- Most arrivals in **August** and **September**
- Waiting list was higher
- Reserved Room Type A – 88%
- People who booked Bed & Breakfast – 77%
- Booked City Hotel – 74%
- Transient customers – 54%
- Distribution Channel – TA/TO = 93%

Interpretation and Results

- To identify potential cancellations, logistic regression or classification tree can be used.
- Customers can be segmented to offer specific incentives to potential cancelers.
- Furthermore, without data analytics, estimating a person's likelihood of canceling a reservation would have been a vague idea.
- To determine cancellation likelihood, the logistic regression was chosen as the final prediction model.
- The performance of logistic regression improves dramatically as the number of the training data grows. Because our training dataset is so enormous, this is most likely what happened in this situation.

Based on validation misclassification rate:

- **Logistic Regression: Best model**
- **K-NN: Worst model**

Managerial Insights

Hotel industry tend to predict cancellations based on advance booking information (aka pick-up models- classical and advanced, where they see that several bookings is picked up from a specific time point to another) for hotel demand forecast.

Classical pick-up model only utilizes completed bookings in their forecasting.

Advanced pick-up method uses both complete and incomplete bookings.

Managers can use our findings in following ways:

- Predict the possibility of cancellations; Incentives can be offered to potential cancellers to attract them to visit the hotels.

- Consider advanced pick-up method (Lead_time being an important variable in our findings)

Some suggestions include:

- Provide a 5% discount, for example, to all possible cancellers.

- On the basis data analysis, a more efficient way would be to offer targeted incentives to potential cancellers, some possible incentives can be provided, keeping in mind the availability of resources and capital when giving incentives:
  - Attract travelers who book Room A - a 5% discount.
  - $20 discount on BB.
  - Travelers/customers from Portugal or Germany will be offered a 5% discount.