CT051-3-M

Data Management

Assignment Part 2

The Impact of Airline Loyalty Program on Flight Activity & Memberships

Student Name: Victor Hew Xin Kai

TP Number: TP078400

Intake Code: APUMF2310DSBA(DE)(PR)

Programme: MSc in Data Science and Business Analytics

Module Lecturer: Dr. Murugananthan Velayutham

Table of Contents

Abstract

Loyalty campaigns are one of the effective ways to build trust between companies and their customers and boost profitability in any industry including airlines through understanding customers' needs and responding to them with personalized services or products. Therefore, the purpose of this report is to analyse the extent of airline loyalty programs in making an impact on the flight activity and membership history of customers. First, the datasets about customer flight activity and loyalty history were integrated into one before it was subjected to data preprocessing. Exploratory Data Analysis (EDA) was then performed on the integrated dataset through data visualisation and statistical representations such as histogram, bar chart, scatterplot, boxplot, Pearson's correlation, chi-square test, two-sample t-test and ANOVA. Feature engineering techniques such as variable creation, one hot encoding and handling outliers, median imputations, log transformation and data normalisation were then introduced followed by feature engineering. The main results demonstrated that the mean CLV among flyers who were enrolled under the 2018 promotion is significantly greater than the standard method. There is a significant difference in points accumulation and redemption, CLV and overall flight bookings between Star, Aurora and Nova loyalty status.

Keywords: Loyalty program, frequent flyer program, airline, data preprocessing, exploratory data analysis, feature engineering

1.0 The Impact of Airline Loyalty Program on Flight Activity & Membership

The competition between different airlines within the dynamic-natured airline industry is very intense. One of the ways for an airline brand to gain the upper hand over its fierce competitors revolves around customer loyalty programs, also specifically known as frequent flyer programs. Recognizing the evolving preference and need of flyers for a more personalised and cost-effective travel experience, airline companies must win the brand loyalty of flyers as the commodity prize so that the values derived from the flyers can be brought back to the airline companies aside from giving the best services to them, which ultimately striking a balance between these two efforts in maximizing the profitability of the airline brands. Frequent flyer programs typically provide airline companies with a marketing edge by implementing policies related to the providence of different incentives and privileges to frequent airline passengers depending on membership tiers so that they feel more satisfied and retain their status in the loyalty program, thus increasing the likelihood of staying loyal to the airline brand in terms of focused expenditure. To do this, they often need to accumulate points through different provided means such as accumulation of flight distance and number of flight reservations and redeem them to obtain point-corresponding prizes like discount fares, premium lounge access and other exclusive promotions. In this way, the airline companies can derive insights from the flight and membership activities of frequent flyers in different aspects such as their demographic information in addition to flight distance and destination as travel preference indicators, and ultimately formulating different personalised benefits that suit their preference to enhance their airline membership experience, which gains their trust for the exceptional service deliveries. Similarly, airline companies can gain economic benefits from them in return as the number of flight seat bookings hikes up.

Thus, this report aims to investigate and analyse the relationship between the implementation of airline loyalty campaigns and flight activities and membership history. First, related works are outlined. Two datasets containing data about the flight activities and membership history of Canadian flyers for one Canada-based airline are then pre-processed to manage the noisy data before proceeding to the exploratory data analysis (EDA) stage where trends and patterns of data are extensively explored through graphical visualisations and statistical analysis. Then, feature engineering techniques like imputation, outlier removal, log

transformation, one hot encoding and standardisation are applied to the data. Finally, five hypotheses are proposed based on the analysis of the cleaned dataset followed by discussions of the findings.

2.0 Related Work

Related past work regarding the involvement of geography-oriented airline loyalty programs has been carried out extensively. de Jong et al. (2019) analysed the existence of any location-wise differences in one-of-a-kind nation-affiliated airline brand commitment behaviours associated with country borders in the global travelling context using confidential data gathered from the loyalty campaign program of the national affiliated airline in the corresponding domestic nation. Their results indicated that the flight distance accumulation of foreign travellers living near the border paled in comparison with local travellers by 60%, which contributed to the decrease in the loyalty program subscription rate by approximately 70%. Batarlienė & Slavinskaitė (2023) investigated the variable importance of airline loyalty programs among 409 travellers in Lithuanian airports using online surveys, with most of the age population being young adults. The main result suggested that the most important factor was the satisfaction of loyal customers towards the airline's image, in addition to other results indicating that the economic and convenience-related benefits motivated approximately 74% of respondents to join the loyalty program.

The involvement and perception of airline passengers towards the loyalty program have also been examined by past studies. Bravo & Vieira (2019) conducted a systematic review to give an overview of customer loyalty in the airline industry and indicated that flight fares played the most significant role in evaluating the loyalty program from the passengers' lens, along with trust between the loyalty member and the airline company. Limberger et al. (2021) examined the extent of participation of flyers in the airline loyalty campaign among 429 respondents in Brazil. Their findings suggested the moderator role of flyers' participation in the association between loyalty rewards and the perceived value of the rewards. Gao et al. (2018) performed further studies on hedonic rewards by investigating the difference between the perceived value of extra and exclusive airline loyalty rewards among 136 online respondents. They found that loyalty status indirectly determined the amount of travelling and had significant direct effects on the perceived value of both rewards and their willingness

to pay, especially for long flight journeys. Research by Lim et al. (2023) focused on studying transactions using loyalty points or real money as the main ‘currency’ for flight reservations using data gathered from a USA-based airline and found that travellers decided their payment method based on how they perceived the points’ value and cross-checked between money and the corresponding points. On the other hand, Orhun et al. (2022) investigated the extent of the involvement of loyalty flyers in climbing up the rank of loyalty status with the decision-making on bearing the flight ticket cost by themselves as the moderator using historical data on customers’ flight activity and loyalty history of a USA-based airline company. They found that flyers who were on the verge of being promoted to a higher loyalty status would be willing to lower their requirements on pricy flight tickets using the provided incentives. Koech et al. (2023) indicated that the perception of travellers towards airline membership campaigns depended on their awareness towards the particular airline and the associated standard.

The prediction of customer lifetime value (CLV) towards the airline loyalty program and flight booking was also done by previous studies. Ruibin & Borglöv (2018) investigated the degree of CLV in examining the performance of the program and found that there was a positive relationship between CLV and the program, which can be facilitated by the perceived pressure of points accumulation and cross-purchasing. However, Thirumuruganathan et al. (2021) wanted to generalise the high prediction accuracy of the developed model (89%) with the use of 2-year data containing over 27 million flight reservations in a USA-based airline company to a real-life context, but to no avail, as the accuracy was 51% only because of the evolving customer preferences towards flights bookings.

3.0 Method

3.1 Data Preprocessing

Using SAS, dimensionality reduction was first carried out by dropping year and month columns in the uploaded flyers’ flight activity dataset, both of which are not necessary. Besides, in the flight activity dataset, the summation of the remaining numerical columns that are additive based on flyers’ loyalty numbers was only possible and logical with the reduction of the year and month column so that the total number of flight reservations as well as point

accumulation and redemption by one unique customer can be calculated. The two datasets revolving around the flight activity and loyalty history of flyers were then integrated into one named as “WORK.Flight_Loyalty” through inner join as both datasets shared the loyalty number of flyers as the common column. After that, the first 3,000 rows of data were selected from the merged dataset for further feature engineering and exploratory data analysis (EDA), and this dataset was now called “WORK.Flight_Loyalty_Extracted”, in which the columns such as loyalty numbers, country and postal code were dropped as they served no meaning in understanding the data, in addition to the country column contained Canada as the only level.

Flight Activity & Loyalty History of Canadian Airline Customers

Obs	Flights Booked	Flights with Companions	Total Flights	Distance	Points Accumulated	Points Redeemed	Dollar Cost Points Redeemed	Province	City	Gender	Education	Salary	Marital Status	Loyalty Card	CLV	Enrollment Type	Enrollment Year	Enrollment Month	Cancellation Year	Cancellation Month
1	132	39	171	51877	5224.44	1418	115	Ontario	Toronto	Female	Bachelor	83236	Married	Star	3839.14	Standard	2016	2	.	.
2	190	25	215	41578	4176.04	1971	159	Alberta	Edmonton	Male	College	.	Divorced	Star	3839.61	Standard	2016	3	.	.
3	66	21	87	19664	1963.00	374	30	British Columbia	Vancouver	Male	College	.	Single	Star	3839.75	Standard	2014	7	2018	1
4	123	36	159	36043	3626.68	1291	105	Ontario	Toronto	Male	College	.	Single	Star	3839.75	Standard	2013	2	.	.
5	132	44	176	36840	3689.68	0	0	Quebec	Hull	Male	Bachelor	103495	Married	Star	3842.79	Standard	2014	10	.	.
6	292	54	346	83996	8464.16	1222	99	Yukon	Whitehorse	Male	Bachelor	51124	Married	Star	3844.57	Standard	2012	5	.	.
7	143	25	168	48292	4880.80	1583	128	Ontario	Toronto	Female	College	.	Single	Star	3857.95	Standard	2014	6	.	.
8	144	41	185	41278	4177.92	733	59	Ontario	Trenton	Male	Bachelor	100159	Married	Star	3861.49	Standard	2016	12	.	.
9	91	33	124	34878	3529.04	1516	123	Quebec	Montreal	Female	Bachelor	100159	Married	Star	3861.49	Standard	2015	5	.	.
10	149	45	194	33622	3413.24	1683	137	Ontario	Toronto	Male	Bachelor	100159	Married	Star	3861.49	Standard	2016	10	.	.

Figure 1: The first 10 observations of the merged dataset – Flight Activity and Loyalty History of Canadian Airline Customers

3.2 Exploratory Data Analysis (EDA)

The figure below details the descriptive statistics of the merged dataset, which include number of observations, mean, mode, median, standard deviation, variance, sum, range and more.

Descriptive Statistics of Customer FLight Activity & Loyalty History

The MEANS Procedure

Variable	N	Mean	Mode	Std Dev	Variance	Minimum	25th Pctl	Median	75th Pctl	Maximum	Sum	Range	Quartile Range
Flights Booked	3000	100.945	0.000	55.052	3030.745	0.000	61.000	114.000	140.000	322.000	302835.000	322.000	79.000
Flights with Companions	3000	25.296	0.000	15.908	253.064	0.000	13.000	26.000	36.000	96.000	75889.000	96.000	23.000
Total Flights	3000	126.241	0.000	68.854	4740.877	0.000	75.000	142.000	175.000	400.000	378724.000	400.000	100.000
Distance	3000	29595.121	0.000	15957.851	254653013.31	0.000	17980.000	34143.500	40897.500	99412.000	88785363.000	99412.000	22917.500
Points Accumulated	3000	3083.199	0.000	1665.574	2774138.026	0.000	1873.750	3531.750	4266.500	10587.500	9249598.340	10587.500	2392.750
Points Redeemed	3000	762.364	0.000	737.377	543724.624	0.000	0.000	582.500	1211.000	4221.000	2287092.000	4221.000	1211.000
Dollar Cost Points Redeemed	3000	61.696	0.000	59.642	3557.185	0.000	0.000	47.000	98.000	343.000	185089.000	343.000	98.000
Salary	2239	80068.889	51573.000	35706.081	1274924232.4	-49830.000	59766.000	74173.000	89645.000	299953.000	179274243.00	349783.000	29879.000
CLV	3000	6378.303	4334.060	2059.960	4243436.861	2004.350	4931.085	5878.690	7752.395	38410.600	19134909.320	36406.250	2821.310
Enrollment Year	3000	2015.249	2018.000	1.972	3.889	2012.000	2014.000	2015.000	2017.000	2018.000	6045748.000	6.000	3.000
Enrollment Month	3000	6.704	7.000	3.379	11.416	1.000	4.000	7.000	10.000	12.000	20113.000	11.000	6.000
Cancellation Year	379	2016.541	2018.000	1.397	1.953	2013.000	2016.000	2017.000	2018.000	2018.000	764269.000	5.000	2.000
Cancellation Month	379	7.061	8.000	3.439	11.830	1.000	4.000	8.000	10.000	12.000	2676.000	11.000	6.000

Figure 2: Descriptive Statistics of Customer Flight Activity and Loyalty History

In terms of univariate analysis, the distributions of numerical variables in the merged dataset using histogram are shown in the figures below. Most of the distributions were positively

skewed except for the enrolment and cancellation year and month distributions, which were approximately uniform.

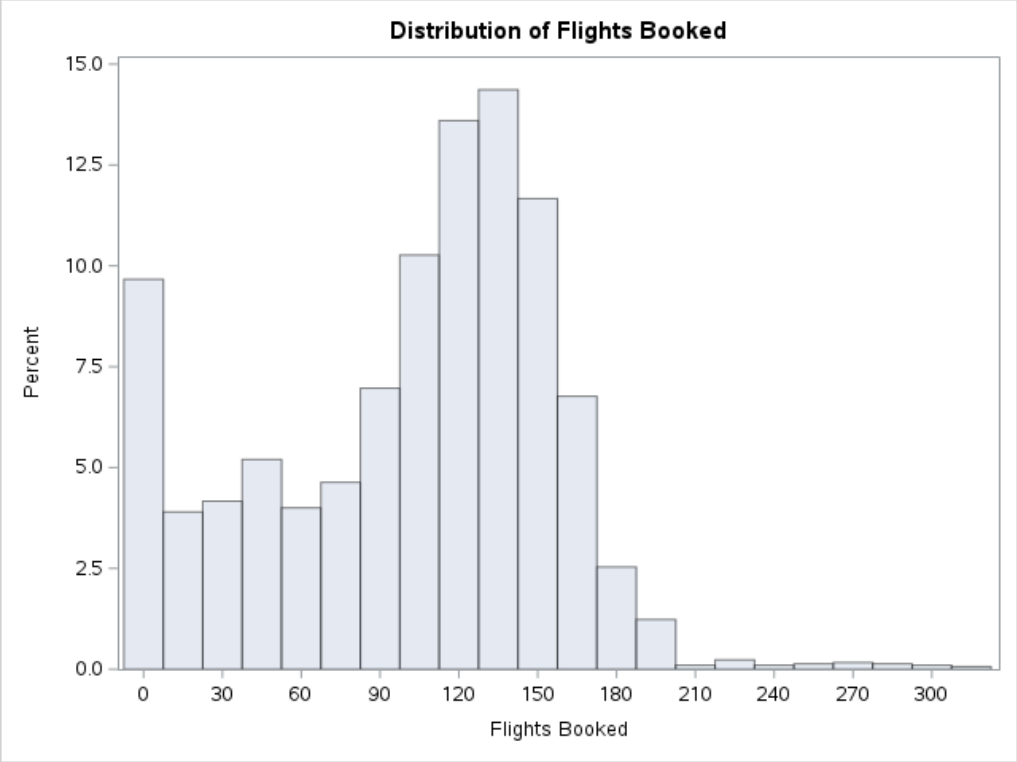


Figure 3: Distribution of Flights Booked

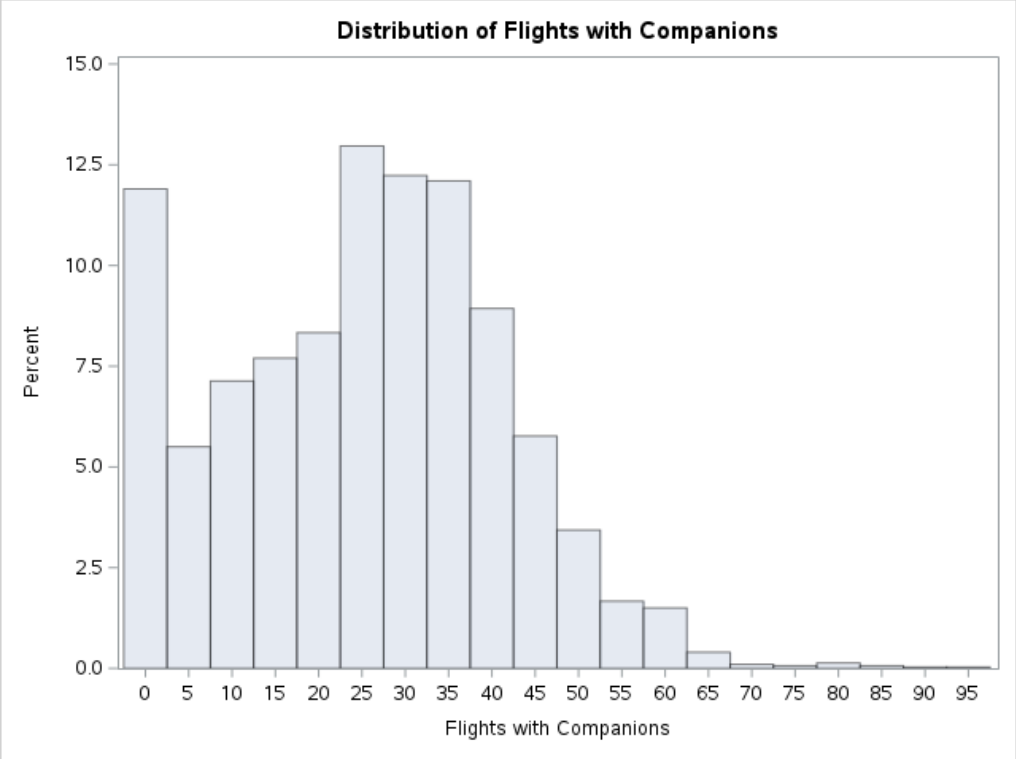


Figure 4: Distribution of Flights with Companions

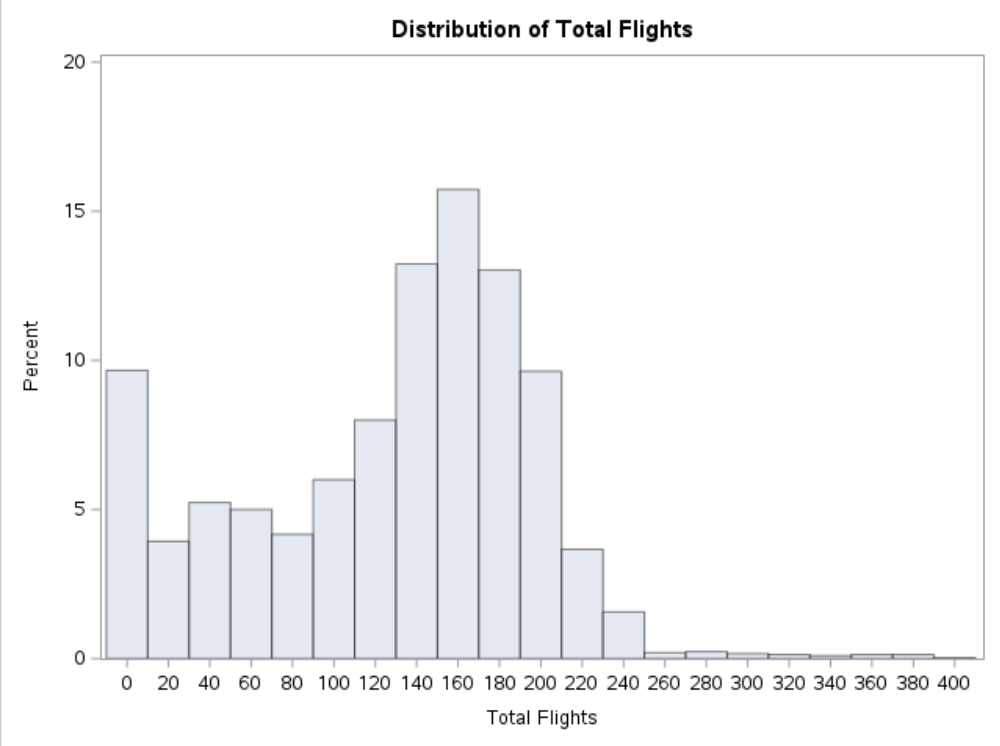


Figure 5: Distribution of Total Flights

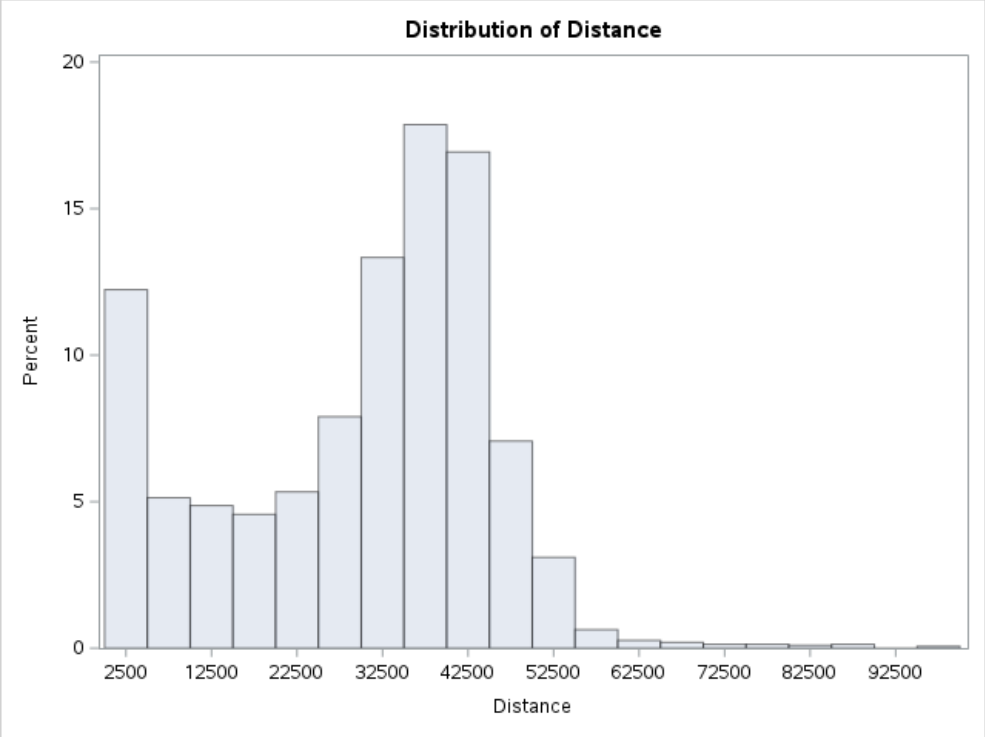


Figure 6: Distribution of Distance

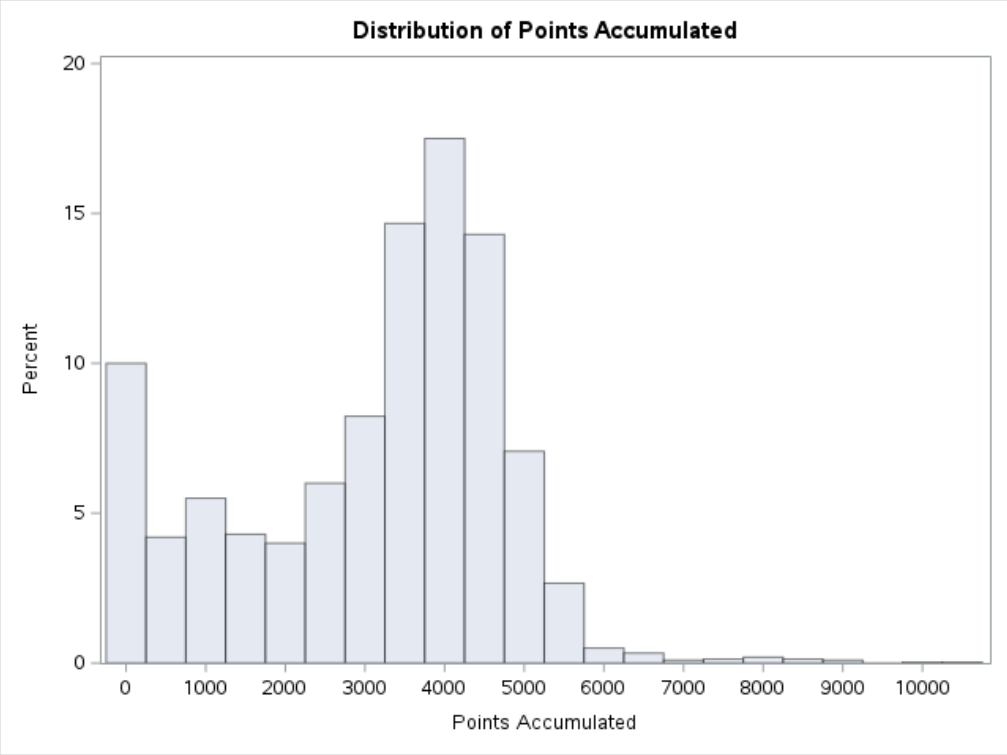


Figure 7: Distribution of Points Accumulated

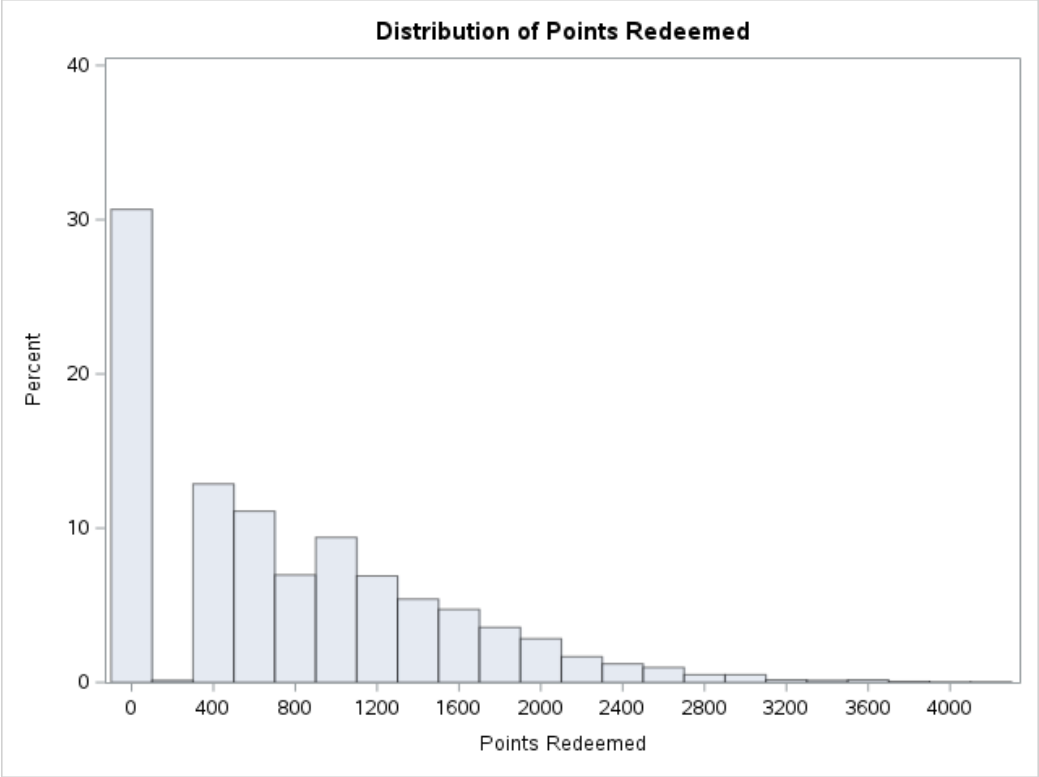


Figure 8: Distribution of Points Redeemed

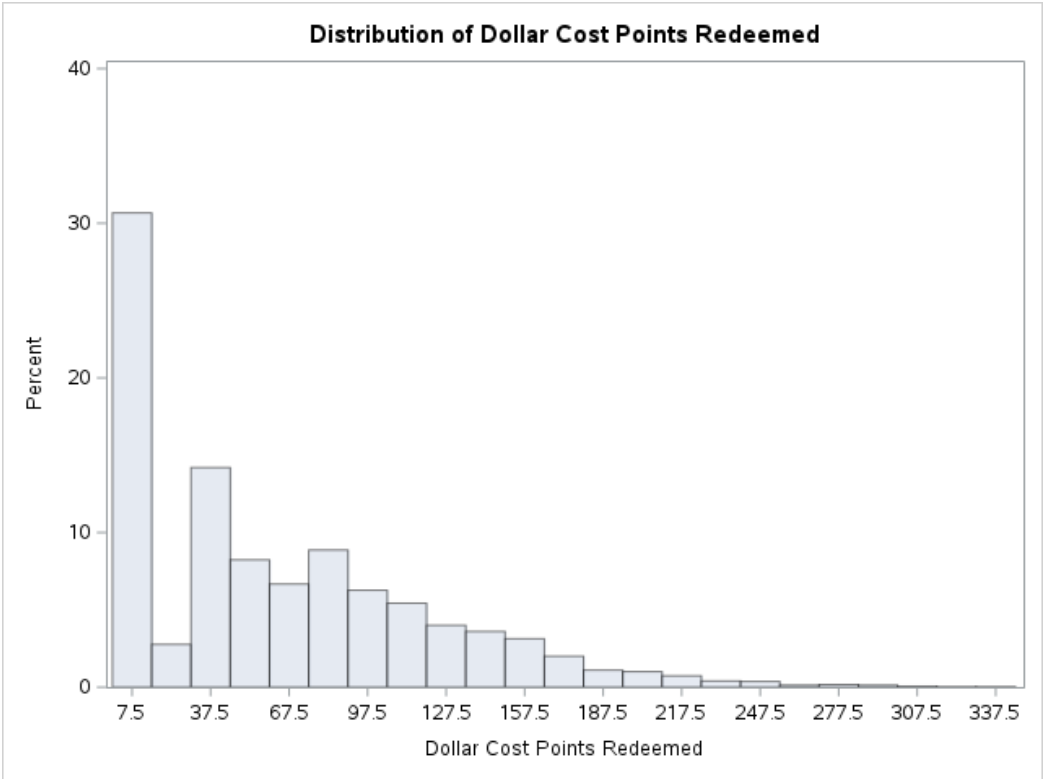


Figure 9: Distribution of Dollar Cost Points Redeemed

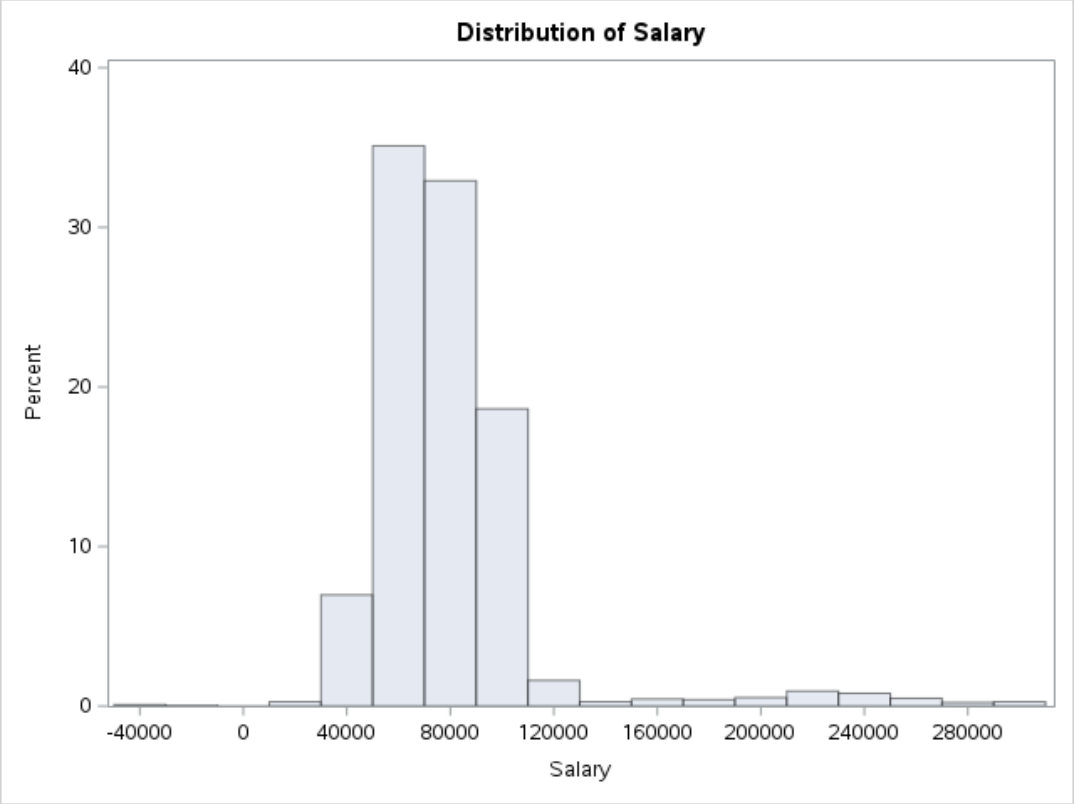


Figure 10: Distribution of Salary

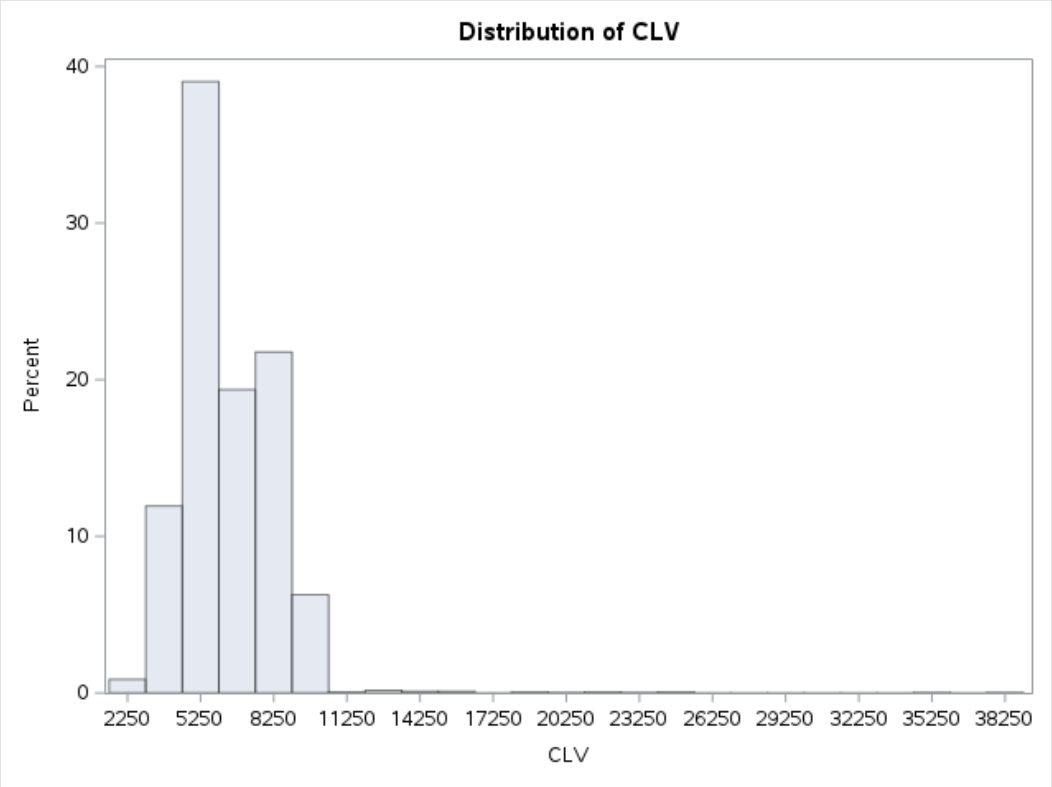


Figure 11: Distribution of CLV

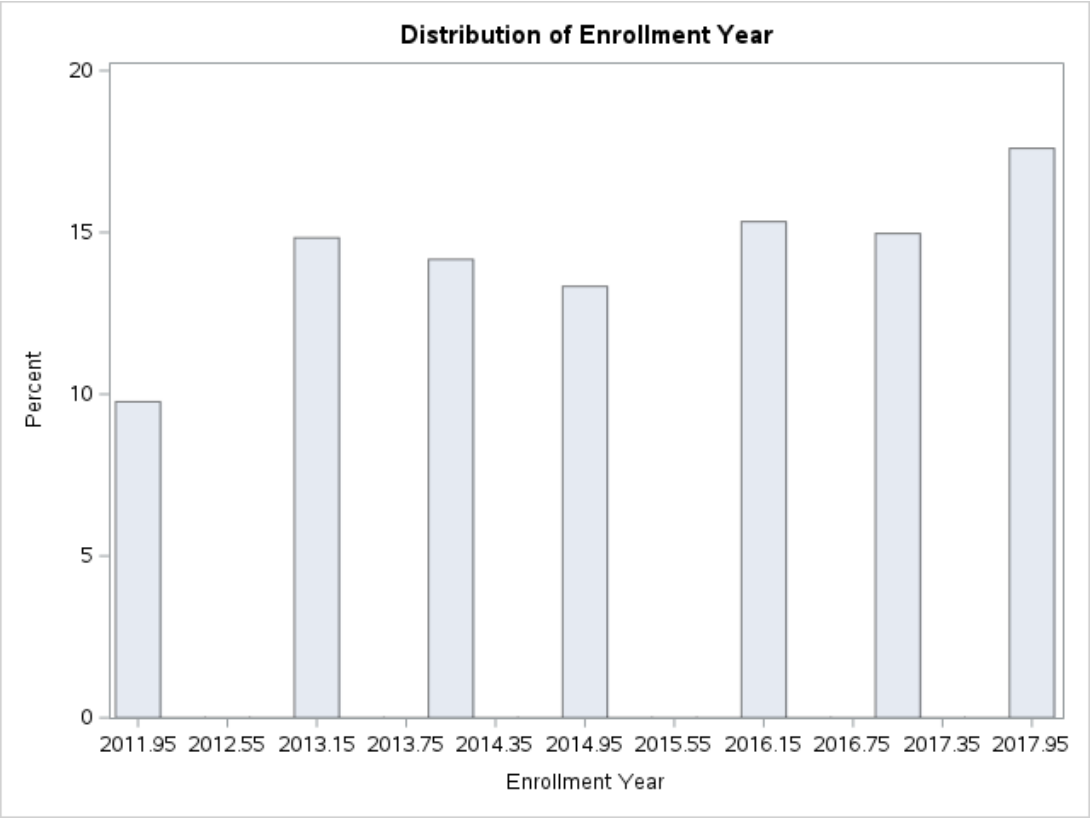


Figure 12: Distribution of Enrollment Year

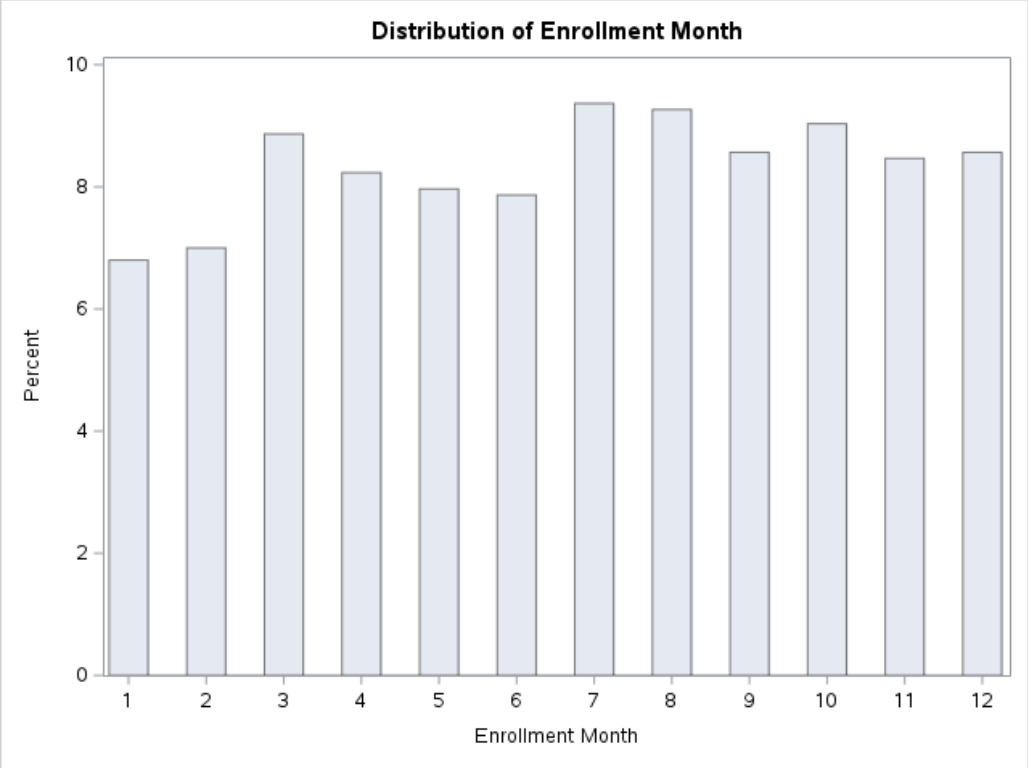


Figure 13: Distribution of Enrollment Month

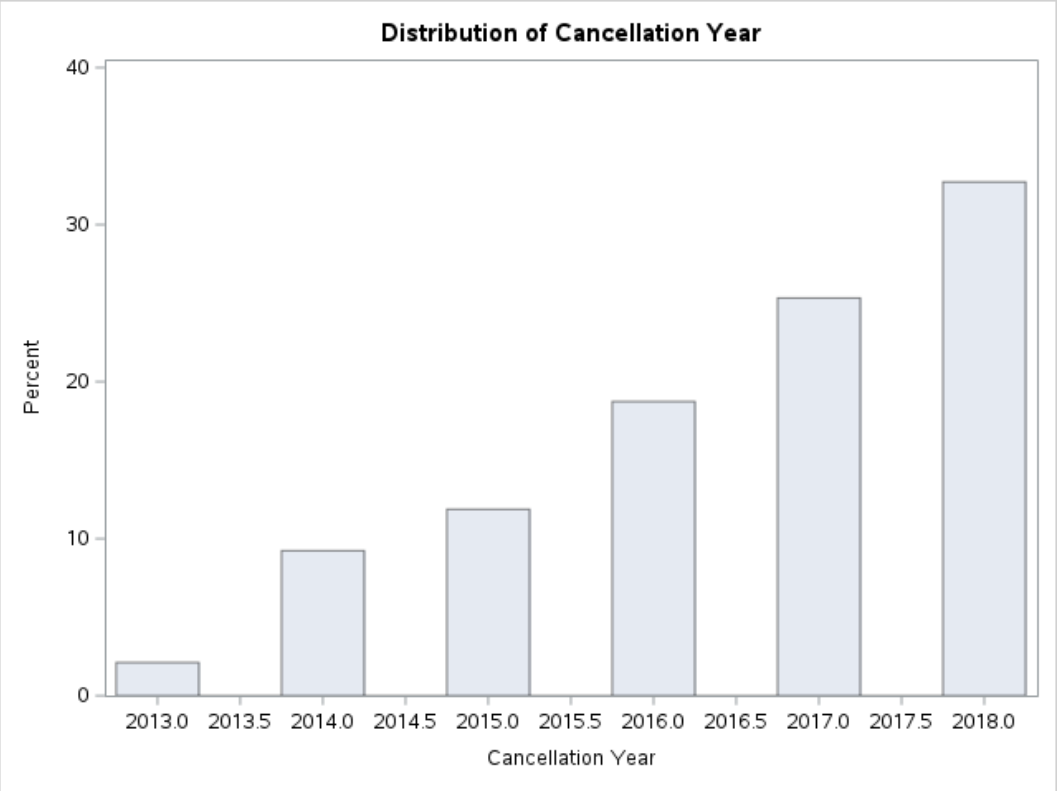


Figure 14: Distribution of Cancellation Year

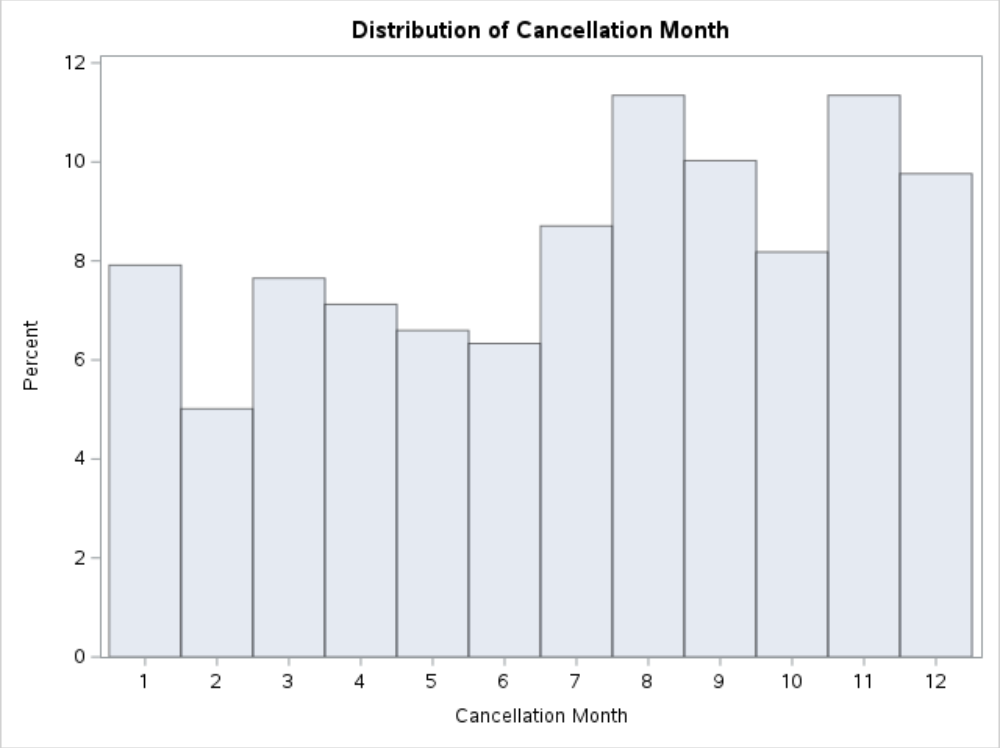


Figure 15: Distribution of Cancellation Month

The frequencies of non-numeric variables were displayed using horizontal bar charts, which are shown in the figures below. Most Canadian customers live in Ontario.

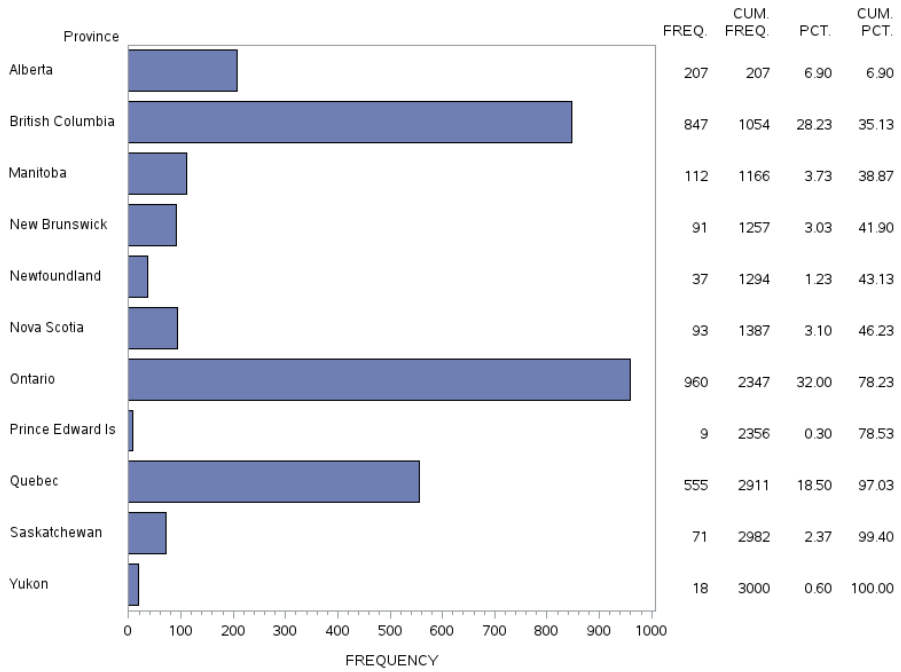
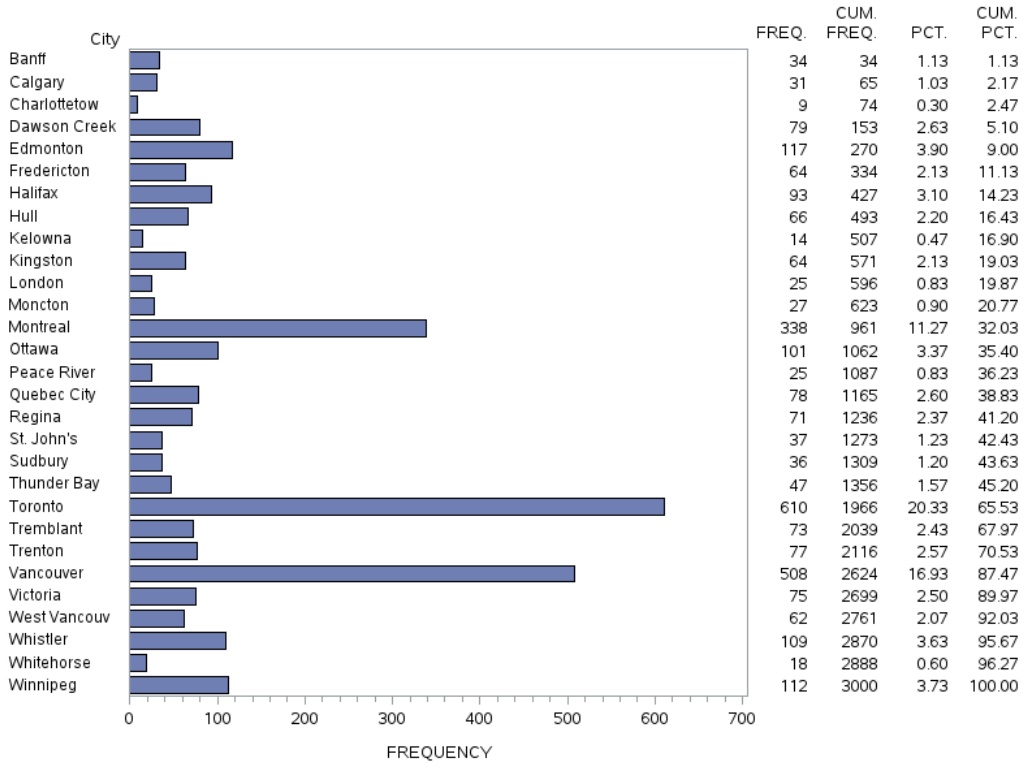


Figure 16: Bar chart for the province

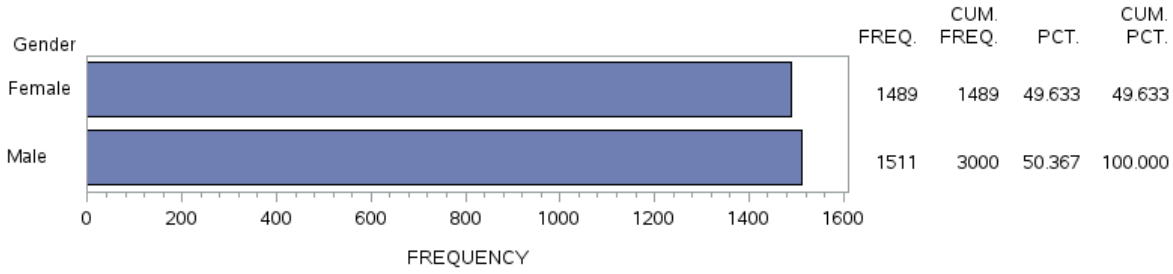
In terms of city, most customers live in Toronto, followed by Vancouver and Montreal.



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Banff	34	34	1.13	1.13
Calgary	31	65	1.03	2.17
Charlottetow	9	74	0.30	2.47
Dawson Creek	79	153	2.63	5.10
Edmonton	117	270	3.90	9.00
Fredericton	64	334	2.13	11.13
Halifax	93	427	3.10	14.23
Hull	66	493	2.20	16.43
Kelowna	14	507	0.47	16.90
Kingston	64	571	2.13	19.03
London	25	596	0.83	19.87
Moncton	27	623	0.90	20.77
Montreal	338	961	11.27	32.03
Ottawa	101	1062	3.37	35.40
Peace River	25	1087	0.83	36.23
Quebec City	78	1165	2.60	38.83
Regina	71	1236	2.37	41.20
St. John's	37	1273	1.23	42.43
Sudbury	36	1309	1.20	43.63
Thunder Bay	47	1356	1.57	45.20
Toronto	610	1966	20.33	65.53
Tremblant	73	2039	2.43	67.97
Trenton	77	2116	2.57	70.53
Vancouver	508	2624	16.93	87.47
Victoria	75	2699	2.50	89.97
West Vancouv	62	2761	2.07	92.03
Whistler	109	2870	3.63	95.67
Whitehorse	18	2888	0.60	96.27
Winnipeg	112	3000	3.73	100.00

Figure 17: Bar chart for city

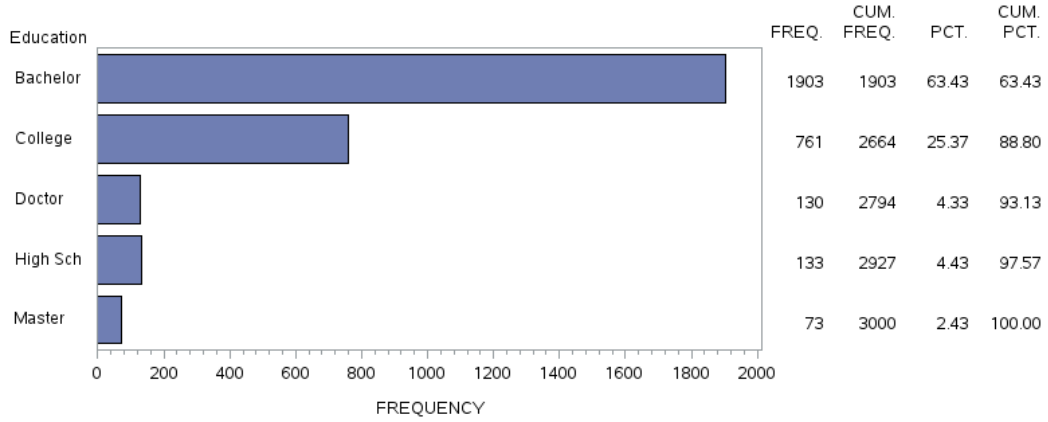
There was no gender imbalance in this merged dataset.



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Female	1489	1489	49.633	49.633
Male	1511	3000	50.367	100.000

Figure 18: Bar chart for gender

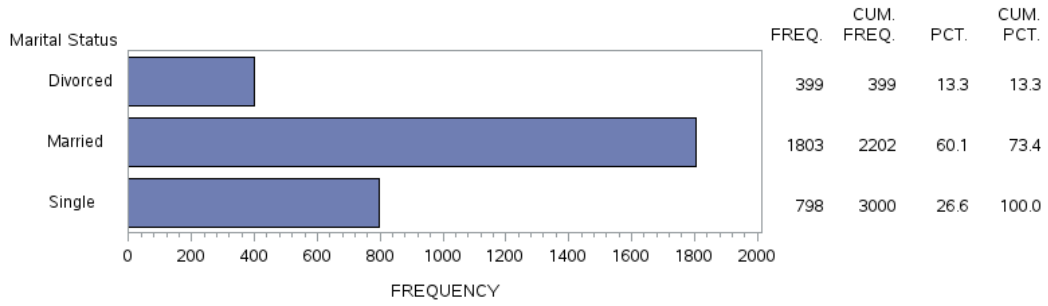
Most customers acquired a bachelor's degree, with the least being a master’s degree.



Education	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Bachelor	1903	1903	63.43	63.43
College	761	2664	25.37	88.80
Doctor	130	2794	4.33	93.13
High Sch	133	2927	4.43	97.57
Master	73	3000	2.43	100.00

Figure 19: Bar chart for educational level

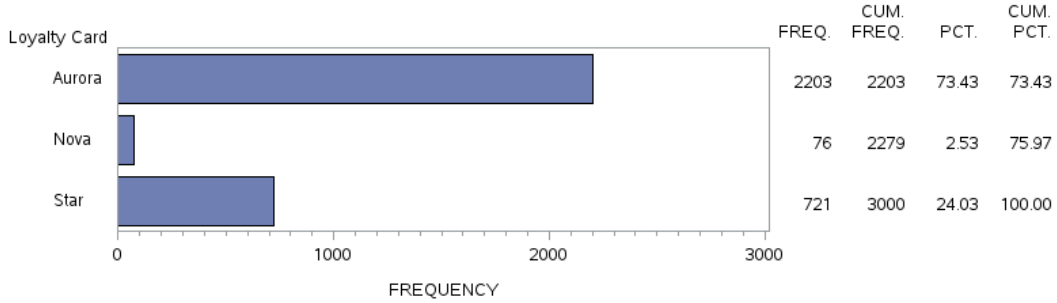
Most flyers were married (60.1%), followed by single (26.6%) and divorced (13.3%).



Marital Status	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Divorced	399	399	13.3	13.3
Married	1803	2202	60.1	73.4
Single	798	3000	26.6	100.0

Figure 20: Bar chart for marital status

Impressively, most flyers who joined the loyalty program possessed the highest card status, which was Aurora.



Loyalty Card	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Aurora	2203	2203	73.43	73.43
Nova	76	2279	2.53	75.97
Star	721	3000	24.03	100.00

Figure 21: Bar chart for loyalty card status

Most Canadian flyers were enrolled in the loyalty program through standard means.

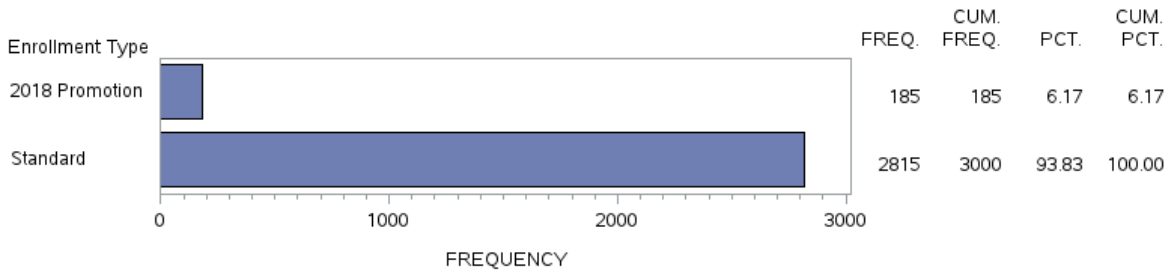


Figure 22: Bar chart for enrolment type

The correlational analysis showed that there was a moderate positive correlation between flight distance and other two variables – points redeemed and the corresponding dollar cost. There was also a strong positive correlation between distance and flight reservations and points accumulated. Moderate positive correlations were found between dollar cost points redeemed and flights booked and with companions, total flights and points accumulated. There was a strong positive correlation between flights booked and points accumulated, but a negligible correlation between flights booked and CLV and the salary of flyers. However, there was a negligible correlation between CLV and salary and the remaining variables. These correlations could be visualised using scatter plots with regression lines, some of which would be displayed below.

The CORR Procedure

9 Variables: CLV Distance Dollar Cost Points Redeemed Flights Booked Flights with Companions Points Accumulated Points Redeemed Salary Total Flights

Pearson Correlation Coefficients
Number of Observations

	CLV	Distance	Dollar Cost Points Redeemed	Flights Booked	Flights with Companions	Points Accumulated	Points Redeemed	Salary	Total Flights
CLV	1.00000 3000	-0.01007 3000	-0.02514 3000	-0.00584 3000	-0.01303 3000	0.00815 3000	-0.02513 3000	0.00113 2239	-0.00768 3000
Distance	-0.01007 3000	1.00000 3000	0.53946 3000	0.95234 3000	0.83852 3000	0.99796 3000	0.53920 3000	0.04290 2239	0.95518 3000
Dollar Cost Points Redeemed	-0.02514 3000	0.53946 3000	1.00000 3000	0.52774 3000	0.57578 3000	0.53722 3000	0.99998 3000	0.00963 2239	0.55498 3000
Flights Booked	-0.00584 3000	0.95234 3000	0.52774 3000	1.00000 3000	0.83188 3000	0.94927 3000	0.52751 3000	0.03349 2239	0.99175 3000
Flights with Companions	-0.01303 3000	0.83852 3000	0.57578 3000	0.83188 3000	1.00000 3000	0.83494 3000	0.57557 3000	0.02921 2239	0.89617 3000
Points Accumulated	0.00815 3000	0.99796 3000	0.53722 3000	0.94927 3000	0.83494 3000	1.00000 3000	0.53696 3000	0.04165 2239	0.95189 3000
Points Redeemed	-0.02513 3000	0.53920 3000	0.99998 3000	0.52751 3000	0.57557 3000	0.53696 3000	1.00000 3000	0.00967 2239	0.55475 3000
Salary	0.00113 2239	0.04290 2239	0.00963 2239	0.03349 2239	0.02921 2239	0.04165 2239	0.00967 2239	1.00000 2239	0.03353 2239
Total Flights	-0.00768 3000	0.95518 3000	0.55498 3000	0.99175 3000	0.89617 3000	0.95189 3000	0.55475 3000	0.03353 2239	1.00000 3000

Figure 23: Correlational analysis

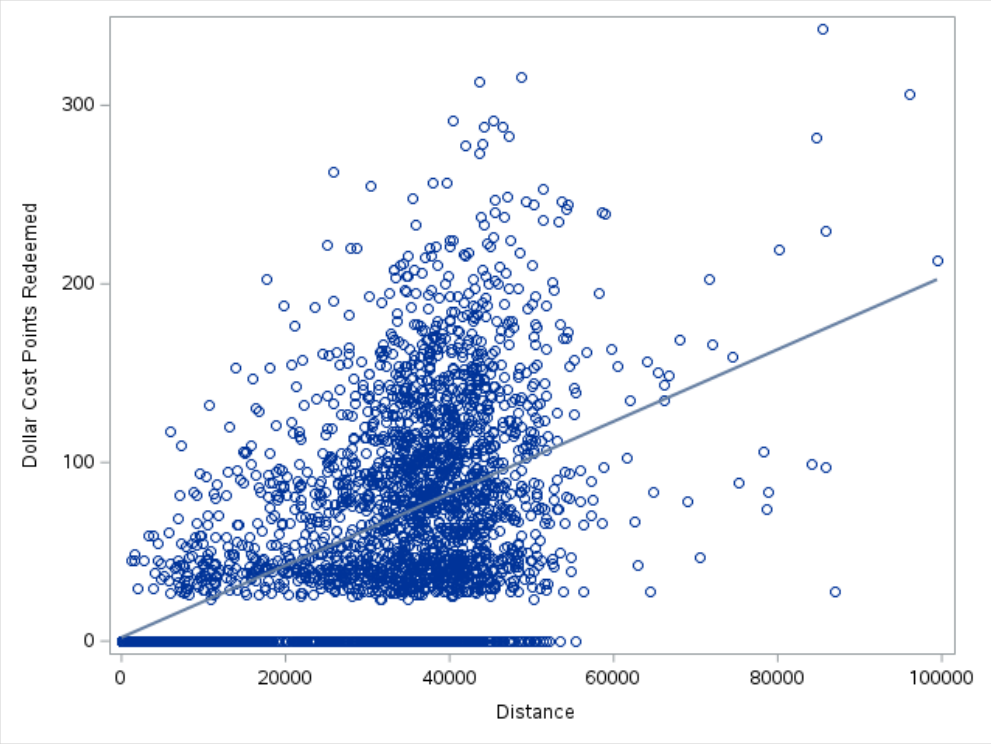


Figure 24: Scatter plot for the relationship between distance and dollar cost points redeemed

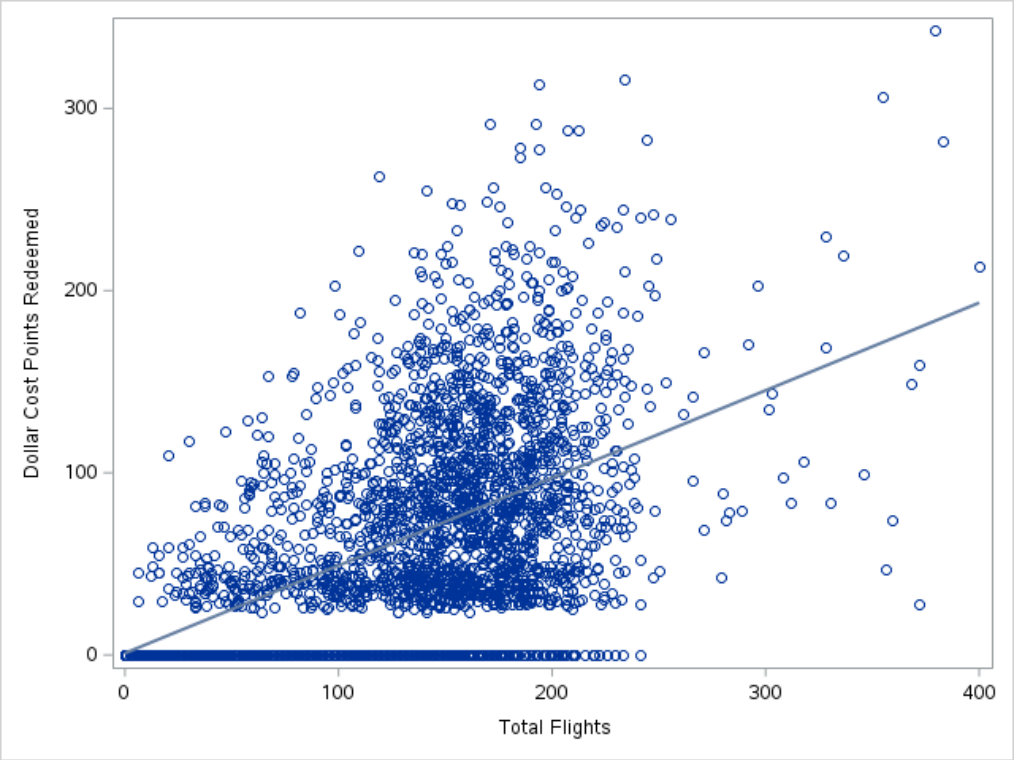


Figure 25: Scatter plot for the relationship between total flights and dollar cost points redeemed

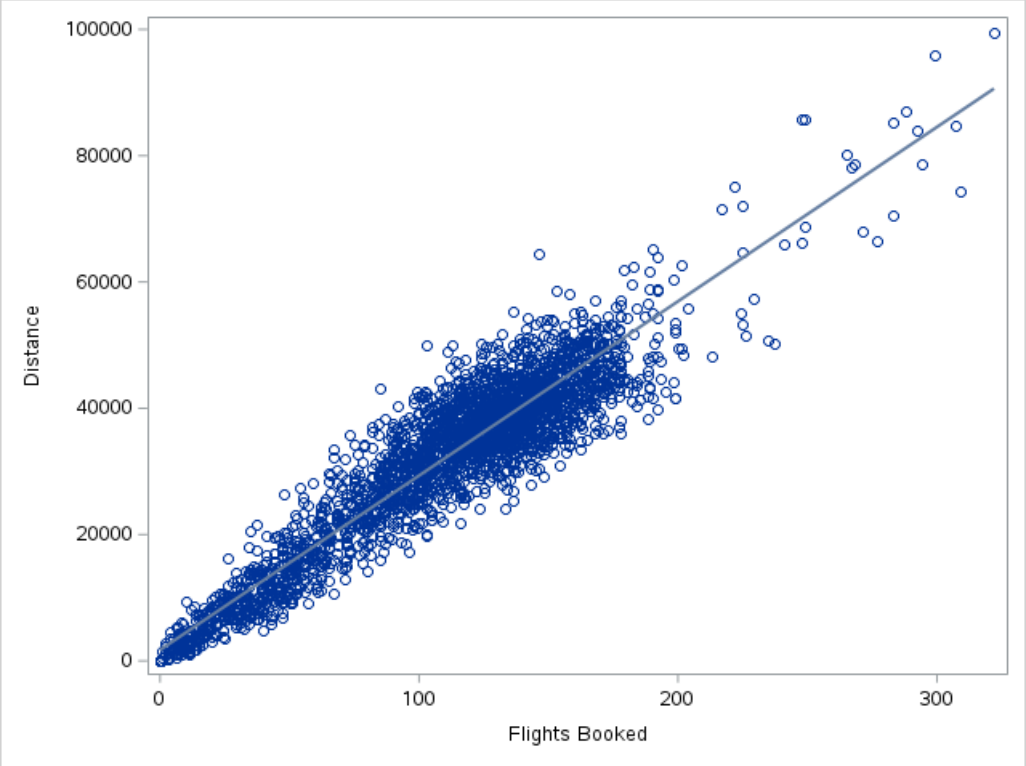


Figure 26: Scatter plot for the relationship between flights booked and distance

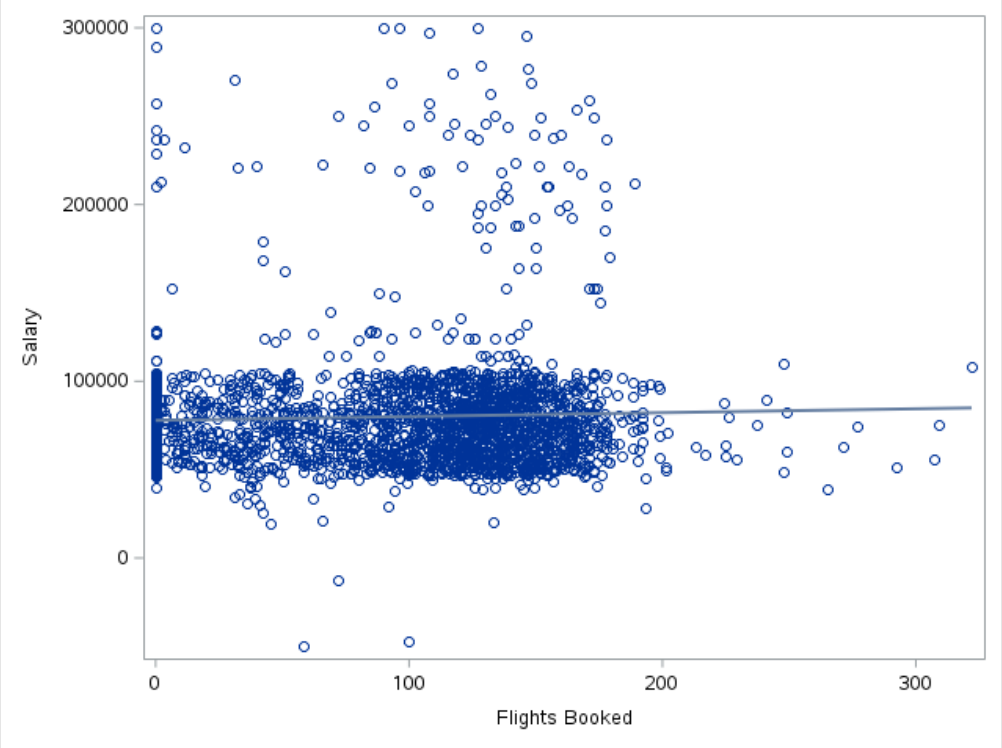


Figure 27: Scatter plot for the relationship between flights booked and salary

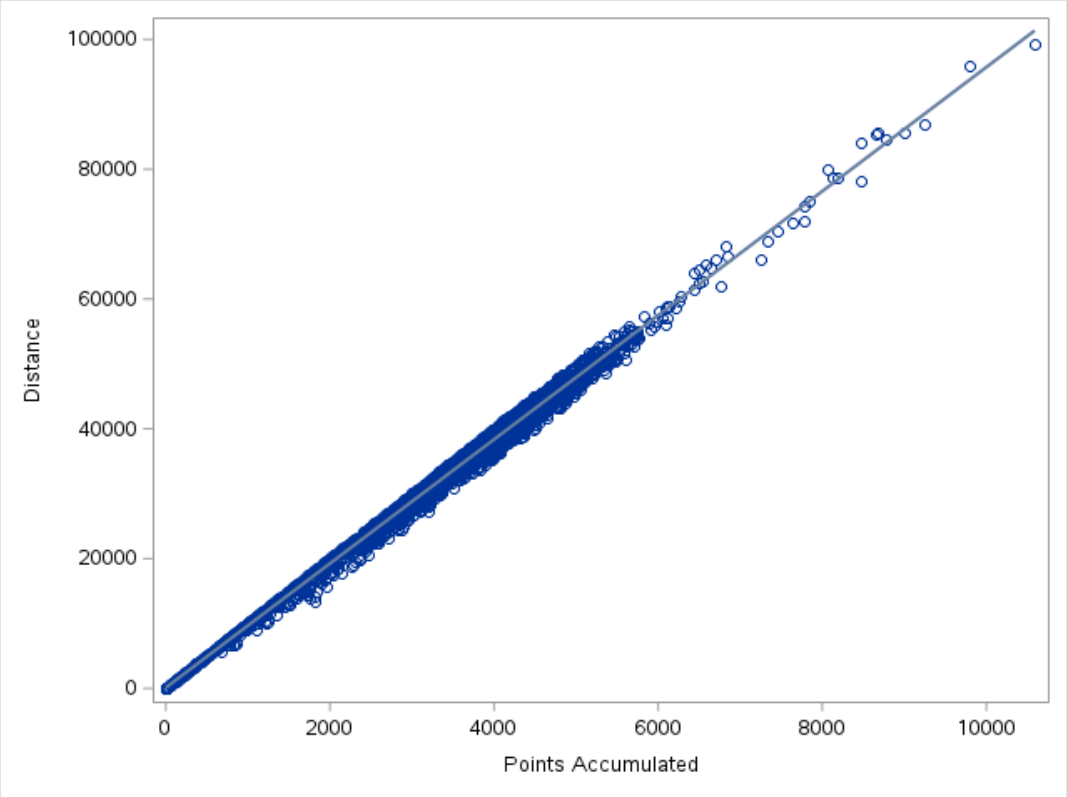


Figure 28: Scatter plot for the relationship between points accumulated and distance

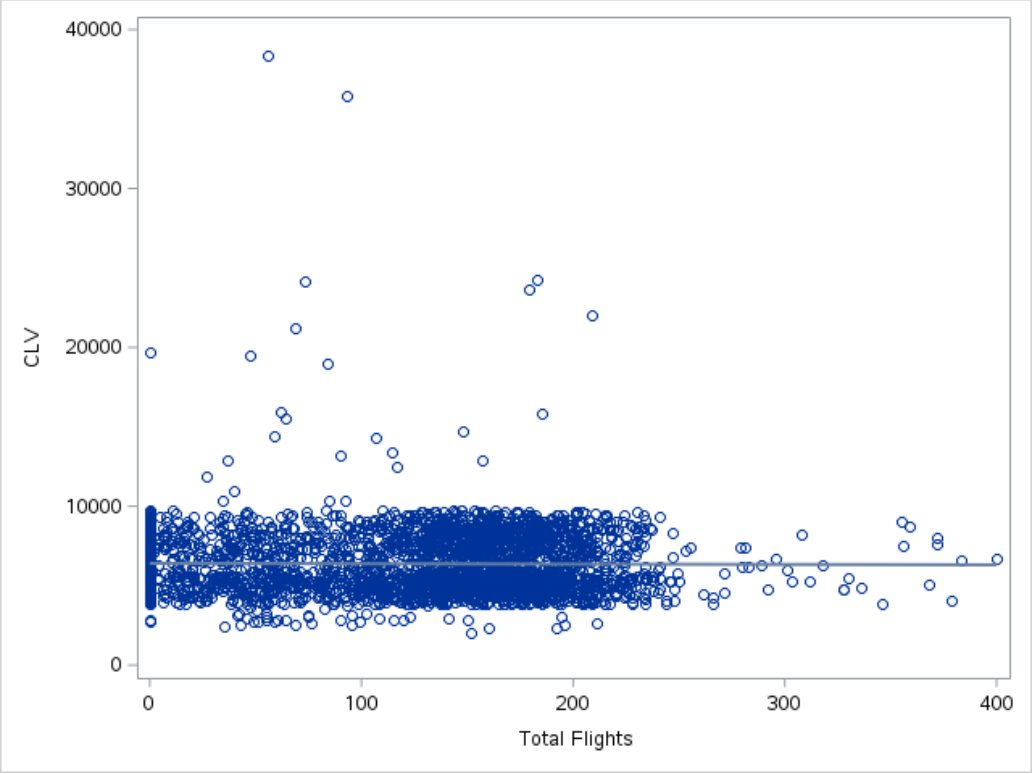


Figure 29: Scatter plot for the relationship between total flights and CLV

Chi-square tests were performed to discover any relationship between two categorical variables. There were statistically significant relationships between loyalty card type and enrolment type, education and marital status respectively, as well as between marital status and enrolment type and education. However, there were no statistically significant relationships between gender and enrolment type, education, loyalty card and marital status.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Gender by Enrollment Type		
	Enrollment Type		
Gender	2018 Promotion	Standard	Total
Female	93 3.10 6.25 50.27	1396 46.53 93.75 49.59	1489 49.63
Male	92 3.07 6.09 49.73	1419 47.30 93.91 50.41	1511 50.37
Total	185 6.17	2815 93.83	3000 100.00

Statistics for Table of Gender by Enrollment Type

Statistic	DF	Value	Prob
Chi-Square	1	0.0320	0.8580
Likelihood Ratio Chi-Square	1	0.0320	0.8580
Continuity Adj. Chi-Square	1	0.0106	0.9180
Mantel-Haenszel Chi-Square	1	0.0320	0.8581
Phi Coefficient		0.0033	
Contingency Coefficient		0.0033	
Cramer's V		0.0033	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	93
Left-sided Pr <= F	0.6006
Right-sided Pr >= F	0.4589
Table Probability (P)	0.0595
Two-sided Pr <= P	0.8795

Sample Size = 3000

Figure 30: Chi-square test on the relationship between gender and enrolment type

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Gender by Education

Gender	Education					
	Bachelor	College	Doctor	High Sch	Master	Total
Female	946 31.53 63.53 49.71	367 12.23 24.65 48.23	71 2.37 4.77 54.62	69 2.30 4.63 51.88	36 1.20 2.42 49.32	1489 49.63
Male	957 31.90 63.34 50.29	394 13.13 26.08 51.77	59 1.97 3.90 45.38	64 2.13 4.24 48.12	37 1.23 2.45 50.68	1511 50.37
Total	1903 63.43	761 25.37	130 4.33	133 4.43	73 2.43	3000 100.00

Statistics for Table of Gender by Education

Statistic	DF	Value	Prob
Chi-Square	4	2.1697	0.7046
Likelihood Ratio Chi-Square	4	2.1714	0.7043
Mantel-Haenszel Chi-Square	1	0.1583	0.6908
Phi Coefficient		0.0269	
Contingency Coefficient		0.0269	
Cramer's V		0.0269	

Sample Size = 3000

Figure 31: Chi-square test on the relationship between gender and education

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Loyalty Card by Education

Loyalty Card	Education					
	Bachelor	College	Doctor	High Sch	Master	Total
Aurora	1487 49.57 67.50 78.14	522 17.40 23.69 68.59	79 2.63 3.59 60.77	76 2.53 3.45 57.14	39 1.30 1.77 53.42	2203 73.43
Nova	49 1.63 64.47 2.57	16 0.53 21.05 2.10	2 0.07 2.63 1.54	6 0.20 7.89 4.51	3 0.10 3.95 4.11	76 2.53
Star	367 12.23 50.90 19.29	223 7.43 30.93 29.30	49 1.63 6.80 37.69	51 1.70 7.07 38.35	31 1.03 4.30 42.47	721 24.03
Total	1903 63.43	761 25.37	130 4.33	133 4.43	73 2.43	3000 100.00

Statistics for Table of Loyalty Card by Education

Statistic	DF	Value	Prob
Chi-Square	8	82.0341	<.0001
Likelihood Ratio Chi-Square	8	78.1173	<.0001
Mantel-Haenszel Chi-Square	1	73.9181	<.0001
Phi Coefficient		0.1654	
Contingency Coefficient		0.1631	
Cramer's V		0.1169	

Sample Size = 3000

Figure 32: Chi-square test on the relationship between loyalty card status and education

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Loyalty Card by Enrollment Type

Loyalty Card	Enrollment Type: 2018 Promotion	Standard	Total
Aurora	70 2.33 3.18 37.84	2133 71.10 96.82 75.77	2203 73.43
Nova	76 2.53 100.00 41.08	0 0.00 0.00 0.00	76 2.53
Star	39 1.30 5.41 21.08	682 22.73 94.59 24.23	721 24.03
Total	185 6.17	2815 93.83	3000 100.00

Statistics for Table of Loyalty Card by Enrollment Type

Statistic	DF	Value	Prob
Chi-Square	2	1191.1656	<.0001
Likelihood Ratio Chi-Square	2	465.1676	<.0001
Mantel-Haenszel Chi-Square	1	28.7447	<.0001
Phi Coefficient		0.6301	
Contingency Coefficient		0.5331	
Cramer's V		0.6301	

Sample Size = 3000

Figure 33: Chi-square test on the relationship between loyalty card type and enrolment type

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Loyalty Card by Marital Status

Loyalty Card	Marital Status: Divorced	Married	Single	Total
Aurora	312 10.40 14.16 78.20	1348 44.93 61.19 74.76	543 18.10 24.65 68.05	2203 73.43
Nova	6 0.20 7.89 1.50	47 1.57 61.84 2.61	23 0.77 30.26 2.88	76 2.53
Star	81 2.70 11.23 20.30	408 13.60 56.59 22.63	232 7.73 32.18 29.07	721 24.03
Total	399 13.30	1803 60.10	798 26.60	3000 100.00

Statistics for Table of Loyalty Card by Marital Status

Statistic	DF	Value	Prob
Chi-Square	4	19.1382	0.0007
Likelihood Ratio Chi-Square	4	19.0861	0.0008
Mantel-Haenszel Chi-Square	1	16.4027	<.0001
Phi Coefficient		0.0799	
Contingency Coefficient		0.0796	
Cramer's V		0.0565	

Sample Size = 3000

Figure 34: Chi-square test on the relationship between loyalty card status and marital status

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Gender by Loyalty Card

Gender	Loyalty Card: Aurora	Nova	Star	Total
Female	1101 36.70 73.94 49.98	35 1.17 2.35 46.05	353 11.77 23.71 48.96	1489 49.63
Male	1102 36.73 72.93 50.02	41 1.37 2.71 53.95	368 12.27 24.35 51.04	1511 50.37
Total	2203 73.43	76 2.53	721 24.03	3000 100.00

Statistics for Table of Gender by Loyalty Card

Statistic	DF	Value	Prob
Chi-Square	2	0.6249	0.7317
Likelihood Ratio Chi-Square	2	0.6254	0.7315
Mantel-Haenszel Chi-Square	1	0.2821	0.5954
Phi Coefficient		0.0144	
Contingency Coefficient		0.0144	
Cramer's V		0.0144	

Sample Size = 3000

Figure 35: Chi-square test on the relationship between gender and loyalty card status

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Enrollment Type by Marital Status

Enrollment Type	Marital Status: Divorced	Married	Single	Total
2018 Promotion	22 0.73 11.89 5.51	109 3.63 58.92 6.05	54 1.80 29.19 6.77	185 6.17
Standard	377 12.57 13.39 94.49	1694 56.47 60.18 93.95	744 24.80 26.43 93.23	2815 93.83
Total	399 13.30	1803 60.10	798 26.60	3000 100.00

Statistics for Table of Enrollment Type by Marital Status

Statistic	DF	Value	Prob
Chi-Square	2	0.8366	0.6582
Likelihood Ratio Chi-Square	2	0.8327	0.6594
Mantel-Haenszel Chi-Square	1	0.8259	0.3635
Phi Coefficient		0.0167	
Contingency Coefficient		0.0167	
Cramer's V		0.0167	

Sample Size = 3000

Figure 36: Chi-square test on the relationship between enrolment type and marital status

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Marital Status by Education

Marital Status	Education					
	Bachelor	College	Doctor	High Sch	Master	Total
Divorced	265 8.83 66.42 13.93	70 2.33 17.54 9.20	30 1.00 7.52 23.08	15 0.50 3.76 11.28	19 0.63 4.76 26.03	399 13.30
Married	1335 44.50 74.04 70.15	257 8.57 14.25 33.77	83 2.77 4.60 63.85	93 3.10 5.16 69.92	35 1.17 1.94 47.95	1803 60.10
Single	303 10.10 37.97 15.92	434 14.47 54.39 57.03	17 0.57 2.13 13.08	25 0.83 3.13 18.80	19 0.63 2.38 26.03	798 26.60
Total	1903 63.43	761 25.37	130 4.33	133 4.43	73 2.43	3000 100.00

Statistics for Table of Marital Status by Education

Statistic	DF	Value	Prob
Chi-Square	8	511.3110	<.0001
Likelihood Ratio Chi-Square	8	474.0898	<.0001
Mantel-Haenszel Chi-Square	1	19.3359	<.0001
Phi Coefficient		0.4128	
Contingency Coefficient		0.3816	
Cramer's V		0.2919	

Sample Size = 3000

Figure 37: Chi-square test on the relationship between marital status and education

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Enrollment Type by Education

Enrollment Type	Education					
	Bachelor	College	Doctor	High Sch	Master	Total
2018 Promotion	127 4.23 68.65 6.67	42 1.40 22.70 5.52	3 0.10 1.62 2.31	10 0.33 5.41 7.52	3 0.10 1.62 4.11	185 6.17
Standard	1776 59.20 63.09 93.33	719 23.97 25.54 94.48	127 4.23 4.51 97.69	123 4.10 4.37 92.48	70 2.33 2.49 95.89	2815 93.83
Total	1903 63.43	761 25.37	130 4.33	133 4.43	73 2.43	3000 100.00

Statistics for Table of Enrollment Type by Education

Statistic	DF	Value	Prob
Chi-Square	4	5.6967	0.2230
Likelihood Ratio Chi-Square	4	6.7300	0.1509
Mantel-Haenszel Chi-Square	1	1.5699	0.2102
Phi Coefficient		0.0436	
Contingency Coefficient		0.0435	
Cramer's V		0.0436	

Sample Size = 3000

Figure 38: Chi-square test on the relationship between enrolment type and education

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Gender by Marital Status			
	Marital Status			
Gender	Divorced	Married	Single	Total
Female	194 6.47 13.03 48.62	893 29.77 59.97 49.53	402 13.40 27.00 50.38	1489 49.63
Male	205 6.83 13.57 51.38	910 30.33 60.23 50.47	396 13.20 26.21 49.62	1511 50.37
Total	399 13.30	1803 60.10	798 26.60	3000 100.00

Statistics for Table of Gender by Marital Status

Statistic	DF	Value	Prob
Chi-Square	2	0.3473	0.8406
Likelihood Ratio Chi-Square	2	0.3474	0.8406
Mantel-Haenszel Chi-Square	1	0.3470	0.5558
Phi Coefficient		0.0108	
Contingency Coefficient		0.0108	
Cramer's V		0.0108	

Sample Size = 3000

Figure 39: Chi-square test on the relationship between gender and marital status

Two-sample t-tests were performed using gender and enrolment type as the class. As the p-value is less than 0.05 significance level, the mean CLV among flyers who were enrolled under the 2018 promotion is significantly greater than the standard method. The mean flight distance, dollar cost points redeemed, points accumulated, points redeemed, salary, total number of flights and number of flights booked with companions among flyers who were under the standard method was significantly greater than the 2018 promotion. However, there were no statistically significant gender differences in all numerical variables.

The TTEST Procedure

Variable: CLV

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	7046.8	5350.5	393.4	2004.4	38410.6
Standard		2815	6334.4	1618.4	30.5031	3839.1	9766.2
Diff (1-2)	Pooled		712.4	2053.2	155.8		
Diff (1-2)	Satterthwaite		712.4		394.6		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		7046.8	6270.7	7822.9	5350.5	4855.2	5959.2
Standard		6334.4	6274.6	6394.2	1618.4	1577.2	1661.8
Diff (1-2)	Pooled	712.4	406.9	1018.0	2053.2	2002.5	2106.5
Diff (1-2)	Satterthwaite	712.4	-65.9623	1490.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	4.57	<.0001
Satterthwaite	Unequal	186.22	1.81	0.0726

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	184	2814	10.93	<.0001

Figure 40: Two-Sample T-test for Differences in CLV Score Between 2018 Promotion and Standard Enrolment

Variable: Distance

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	21015.0	12028.5	884.4	0	48348.0
Standard		2815	30159.0	16024.2	302.0	0	99412.0
Diff (1-2)	Pooled		-9144.0	15808.1	1199.8		
Diff (1-2)	Satterthwaite		-9144.0		934.5		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		21015.0	19270.2	22759.8	12028.5	10915.1	13397.0
Standard		30159.0	29566.8	30751.2	16024.2	15616.3	16454.2
Diff (1-2)	Pooled	-9144.0	-11496.6	-6791.5	15808.1	15418.0	16218.7
Diff (1-2)	Satterthwaite	-9144.0	-10985.3	-7302.7			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-7.62	<.0001
Satterthwaite	Unequal	229.22	-9.78	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.77	<.0001

Figure 41: Two-Sample T-test for Differences in Flight Distance Between 2018 Promotion and Standard Enrolment

Variable: Dollar Cost Points Redeemed

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	41.1676	51.1094	3.7576	0	255.0
Standard		2815	63.0455	59.9226	1.1294	0	343.0
Diff (1-2)	Pooled		-21.8779	59.4193	4.5099		
Diff (1-2)	Satterthwaite		-21.8779		3.9237		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		41.1676	33.7540	48.5812	51.1094	46.3782	56.9240
Standard		63.0455	60.8309	65.2600	59.9226	58.3972	61.5303
Diff (1-2)	Pooled	-21.8779	-30.7206	-13.0352	59.4193	57.9528	60.9626
Diff (1-2)	Satterthwaite	-21.8779	-29.6110	-14.1448			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-4.85	<.0001
Satterthwaite	Unequal	218.63	-5.58	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.37	0.0052

Figure 42: Two-Sample T-test for Differences in Dollar Cost Points Redeemed Between 2018 Promotion and Standard Enrolment

Variable: Flights Booked

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	75.0595	43.4124	3.1917	0	193.0
Standard		2815	102.6	55.3156	1.0426	0	322.0
Diff (1-2)	Pooled		-27.5867	54.6598	4.1486		
Diff (1-2)	Satterthwaite		-27.5867		3.3577		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		75.0595	68.7623	81.3566	43.4124	39.3937	48.3513
Standard		102.6	100.6	104.7	55.3156	53.9076	56.7997
Diff (1-2)	Pooled	-27.5867	-35.7211	-19.4523	54.6598	53.3107	56.0794
Diff (1-2)	Satterthwaite	-27.5867	-34.2033	-20.9702			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-6.65	<.0001
Satterthwaite	Unequal	225.19	-8.22	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.62	<.0001

Figure 43: Two-Sample T-test for Differences in Number of Flights Booked Between 2018 Promotion and Standard Enrolment

Variable: Flights with Companions

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	18.1514	12.4891	0.9182	0	60.0000
Standard		2815	25.7659	15.9977	0.3015	0	96.0000
Diff (1-2)	Pooled		-7.6145	15.8048	1.1996		
Diff (1-2)	Satterthwaite		-7.6145		0.9665		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		18.1514	16.3398	19.9629	12.4891	11.3330	13.9099
Standard		25.7659	25.1747	26.3571	15.9977	15.5904	16.4269
Diff (1-2)	Pooled	-7.6145	-9.9666	-5.2625	15.8048	15.4147	16.2152
Diff (1-2)	Satterthwaite	-7.6145	-9.5190	-5.7101			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-6.35	<.0001
Satterthwaite	Unequal	225.65	-7.88	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.64	<.0001

Figure 44: Two-Sample T-test for Differences in Number of Flights Booked with Companions Between 2018 Promotion and Standard Enrolment

Variable: Points Accumulated

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	2170.0	1254.6	92.2392	0	5214.0
Standard		2815	3143.2	1671.9	31.5112	0	10587.5
Diff (1-2)	Pooled		-973.2	1649.3	125.2		
Diff (1-2)	Satterthwaite		-973.2		97.4732		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		2170.0	1988.0	2352.0	1254.6	1138.5	1397.3
Standard		3143.2	3081.4	3205.0	1671.9	1629.3	1716.7
Diff (1-2)	Pooled	-973.2	-1218.7	-727.8	1649.3	1608.6	1692.1
Diff (1-2)	Satterthwaite	-973.2	-1165.3	-781.1			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-7.77	<.0001
Satterthwaite	Unequal	229.25	-9.98	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.78	<.0001

Figure 45: Two-Sample T-test for Differences in Points Accumulated Between 2018 Promotion and Standard Enrolment

Variable: Points Redeemed

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	508.4	631.8	46.4529	0	3147.0
Standard		2815	779.1	740.8	13.9633	0	4221.0
Diff (1-2)	Pooled		-270.6	734.6	55.7568		
Diff (1-2)	Satterthwaite		-270.6		48.5062		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		508.4	416.8	600.1	631.8	573.3	703.7
Standard		779.1	751.7	806.4	740.8	722.0	760.7
Diff (1-2)	Pooled	-270.6	-379.9	-161.3	734.6	716.5	753.7
Diff (1-2)	Satterthwaite	-270.6	-366.2	-175.0			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-4.85	<.0001
Satterthwaite	Unequal	218.64	-5.58	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.37	0.0052

Figure 46: Two-Sample T-test for Differences in Points Redeemed Between 2018 Promotion and Standard Enrolment

Variable: Salary

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		143	67627.0	35006.8	2927.4	-49830.0	271085
Standard		2096	80917.7	35603.3	777.7	39000.0	299953
Diff (1-2)	Pooled		-13290.7	35565.8	3073.9		
Diff (1-2)	Satterthwaite		-13290.7		3029.0		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		67627.0	61840.1	73414.0	35006.8	31365.7	39611.8
Standard		80917.7	79392.7	82442.8	35603.3	34557.3	36715.2
Diff (1-2)	Pooled	-13290.7	-19318.8	-7262.7	35565.8	34553.5	36639.5
Diff (1-2)	Satterthwaite	-13290.7	-19271.9	-7309.6			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2237	-4.32	<.0001
Satterthwaite	Unequal	162.69	-4.39	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2095	142	1.03	0.8116

Figure 46: Two-Sample T-test for Differences in Salary of Flyers Between 2018 Promotion and Standard Enrolment

Variable: Total Flights

Enrollment Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2018 Promotion		185	93.2108	53.7760	3.9537	0	233.0
Standard		2815	128.4	69.1883	1.3040	0	400.0
Diff (1-2)	Pooled		-35.2013	68.3426	5.1871		
Diff (1-2)	Satterthwaite		-35.2013		4.1632		

Enrollment Type	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2018 Promotion		93.2108	85.4104	101.0	53.7760	48.7980	59.8940
Standard		128.4	125.9	131.0	69.1883	67.4271	71.0446
Diff (1-2)	Pooled	-35.2013	-45.3720	-25.0306	68.3426	66.6558	70.1176
Diff (1-2)	Satterthwaite	-35.2013	-43.4049	-26.9976			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2998	-6.79	<.0001
Satterthwaite	Unequal	226.04	-8.46	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2814	184	1.66	<.0001

Figure 47: Two-Sample T-test for Differences in Total Flights Between 2018 Promotion and Standard Enrolment

One-way ANOVA was also carried out to examine differences in numerical variables among three or more levels of categorical variables. As a result, as the p-value was less than 0.05 alpha level, the mean CLV score and salary were different between the educational levels, however, it was not the same case for points redeemed and its dollar cost counterpart. The mean number of total flights booked, and points accumulated has no difference between the educational levels as the p-value was more than 0.05 alpha level.

The mean CLV score, dollar cost points redeemed, points accumulated, and total flights booked were different between loyalty card types, except for salary. The mean CLV score, dollar cost points redeemed, points accumulated, and total flights have no difference between marital statuses, except for salary.

The ANOVA Procedure

Dependent Variable: CLV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	73277118	18319280	4.34	0.0017
Error	2995	12652790029	4224638		
Corrected Total	2999	12726067147			

R-Square	Coeff Var	Root MSE	CLV Mean
0.005758	32.22475	2055.392	6378.303

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	4	73277118.47	18319279.62	4.34	0.0017

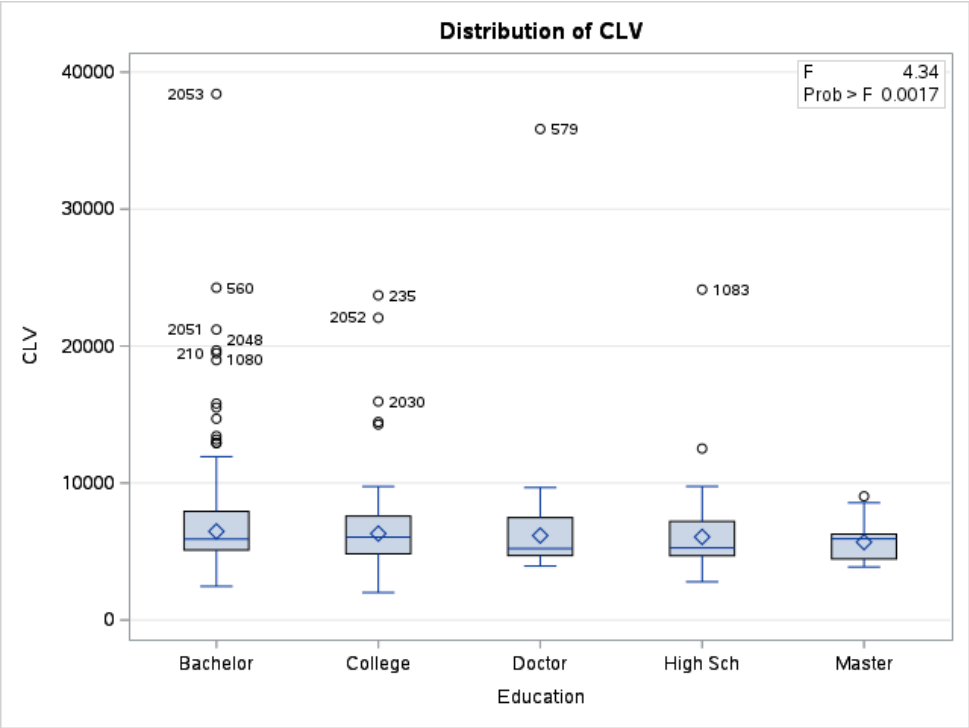


Figure 48: ANOVA Outputs for Differences in CLV Between Educational Levels

Dependent Variable: Dollar Cost Points Redeemed

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	17290.13	4322.53	1.22	0.3020
Error	2995	10650706.23	3556.16		
Corrected Total	2999	10667996.36			

R-Square	Coeff Var	Root MSE	Dollar Cost Points Redeemed Mean
0.001621	96.65658	59.63357	61.69633

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	4	17290.13300	4322.53325	1.22	0.3020

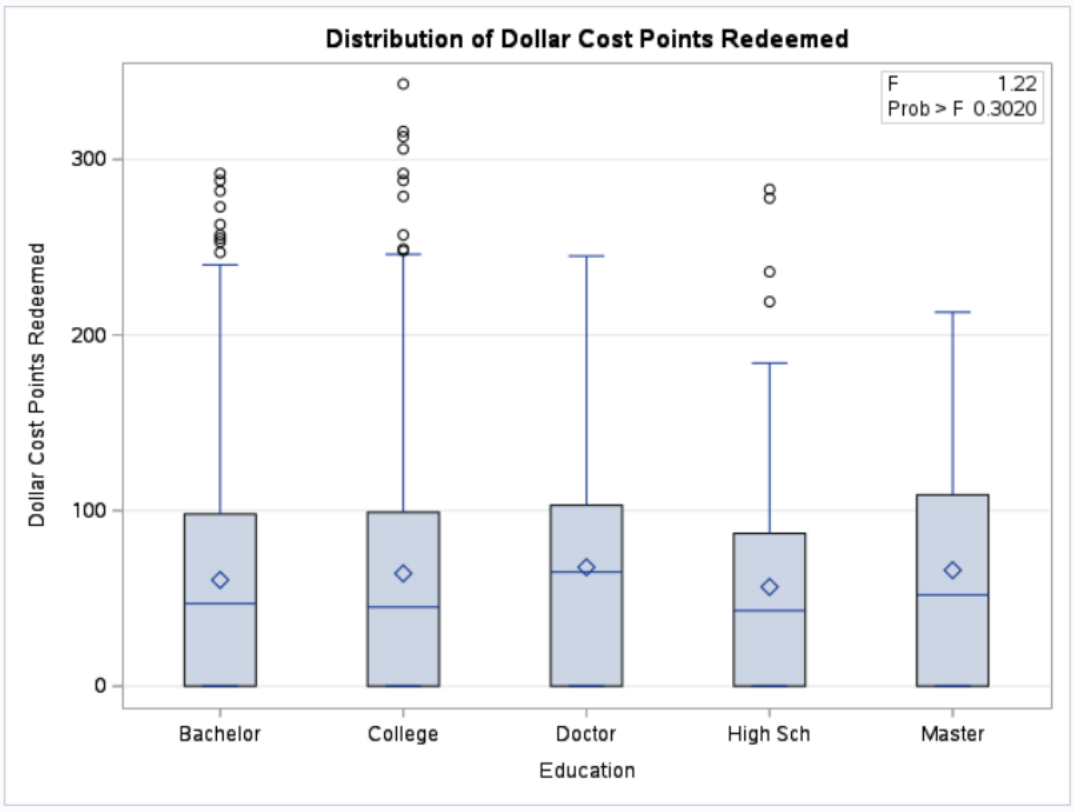


Figure 49: ANOVA Outputs for Differences in Dollar Cost Points Redeemed Between Educational Levels

The ANOVA Procedure

Dependent Variable: Points Accumulated

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	24666946	6166736	2.23	0.0638
Error	2995	8294972994	2769607		
Corrected Total	2999	8319639939			

R-Square	Coeff Var	Root MSE	Points Accumulated Mean
0.002965	53.97684	1664.214	3083.199

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	4	24666945.57	6166736.39	2.23	0.0638

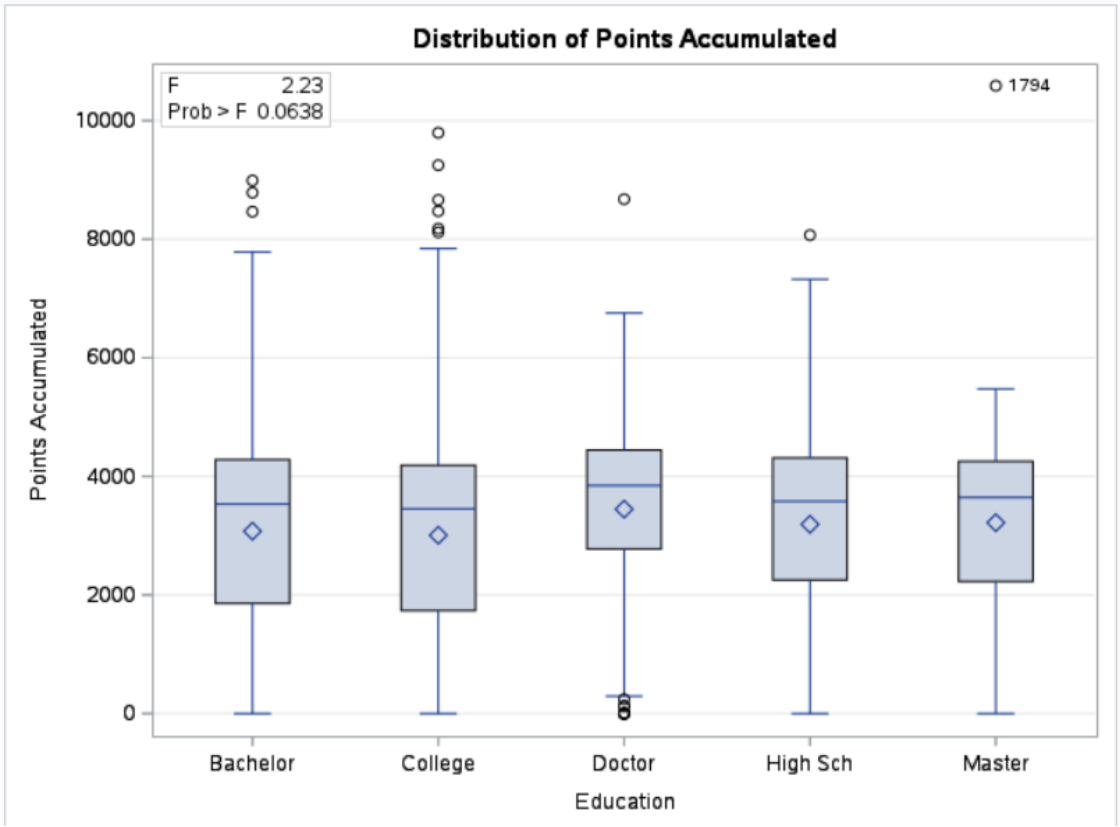


Figure 50: ANOVA Outputs for Differences in Points Accumulated Between Educational Levels

Dependent Variable: Total Flights

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	34254.05	8563.51	1.81	0.1244
Error	2995	14183635.22	4735.77		
Corrected Total	2999	14217889.27			

R-Square	Coeff Var	Root MSE	Total Flights Mean
0.002409	54.51221	68.81694	126.2413

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	4	34254.05406	8563.51351	1.81	0.1244

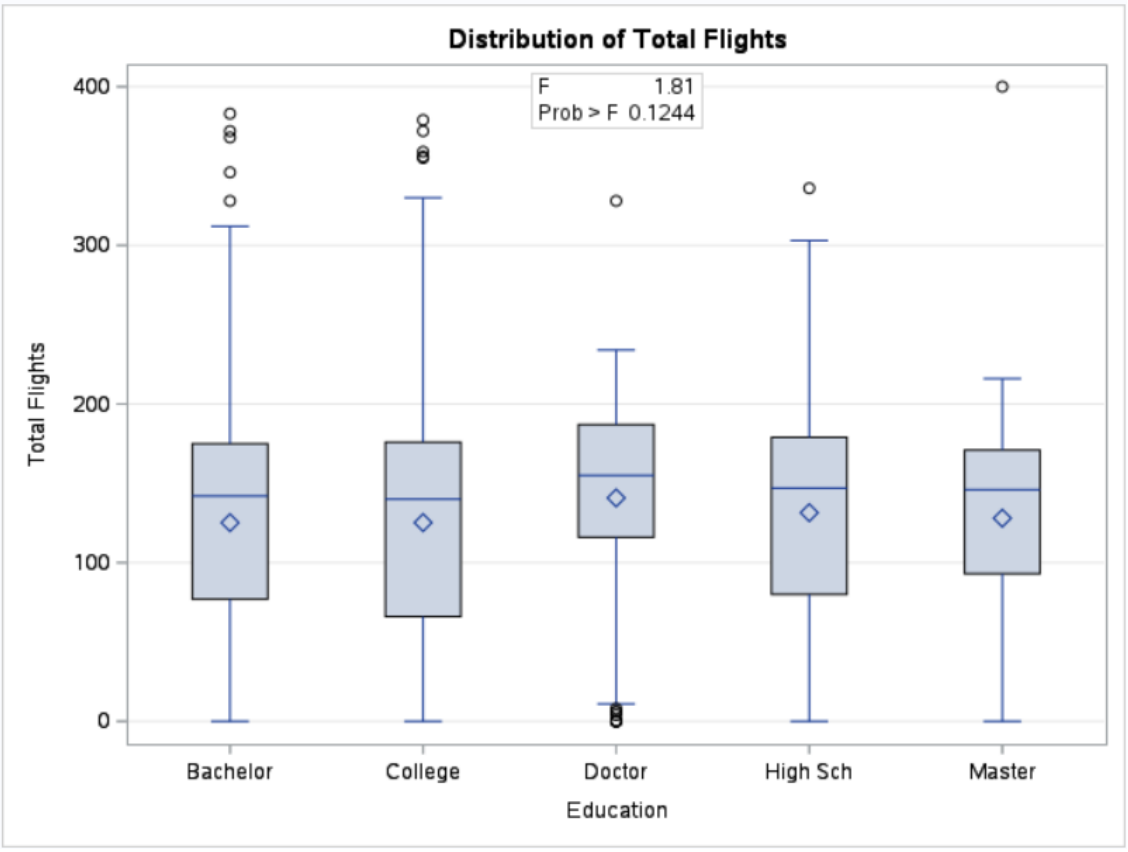


Figure 51: ANOVA Outputs for Differences in Total Flights Between Educational Levels

The ANOVA Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.6210301E12	540343362062	980.05	<.0001
Error	2235	1.2322503E12	551342436.67		
Corrected Total	2238	2.8532804E12			

R-Square	Coeff Var	Root MSE	Salary Mean
0.568129	29.32560	23480.68	80068.89

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Education	3	1.6210301E12	540343362062	980.05	<.0001

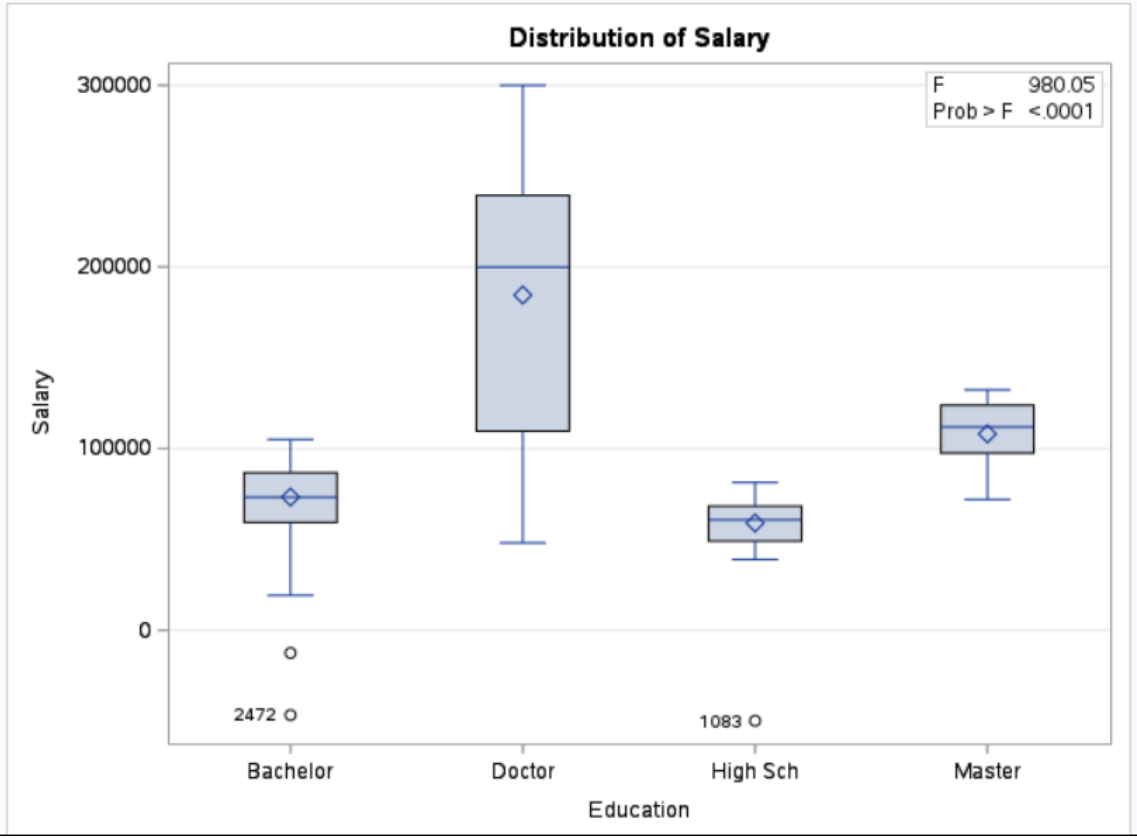


Figure 52: ANOVA Outputs for Differences in Salary Between Educational Levels

The ANOVA Procedure

Dependent Variable: CLV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3389558683	1694779342	544.02	<.0001
Error	2997	9336508464	3115285		
Corrected Total	2999	12726067147			

R-Square	Coeff Var	Root MSE	CLV Mean
0.266348	27.67220	1765.017	6378.303

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Loyalty Card	2	3389558683	1694779342	544.02	<.0001

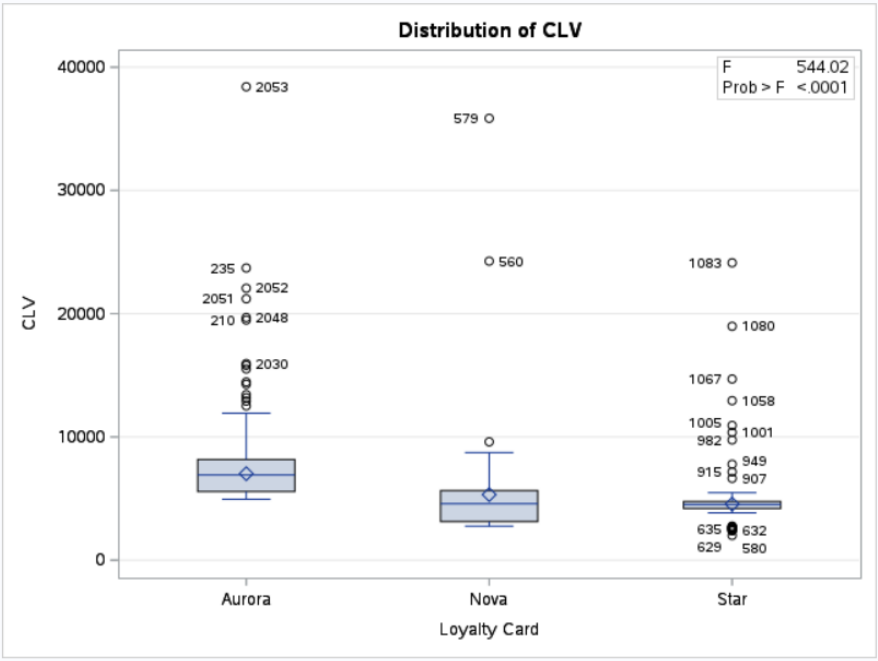


Figure 53: ANOVA Outputs for Differences in CLV Score Between Loyalty Card Type

The ANOVA Procedure

Dependent Variable: Dollar Cost Points Redeemed

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	30227.85	15113.92	4.26	0.0142
Error	2997	10637768.51	3549.47		
Corrected Total	2999	10667996.36			

R-Square	Coeff Var	Root MSE	Dollar Cost Points Redeemed Mean
0.002834	96.56562	59.57745	61.69633

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Loyalty Card	2	30227.84769	15113.92385	4.26	0.0142

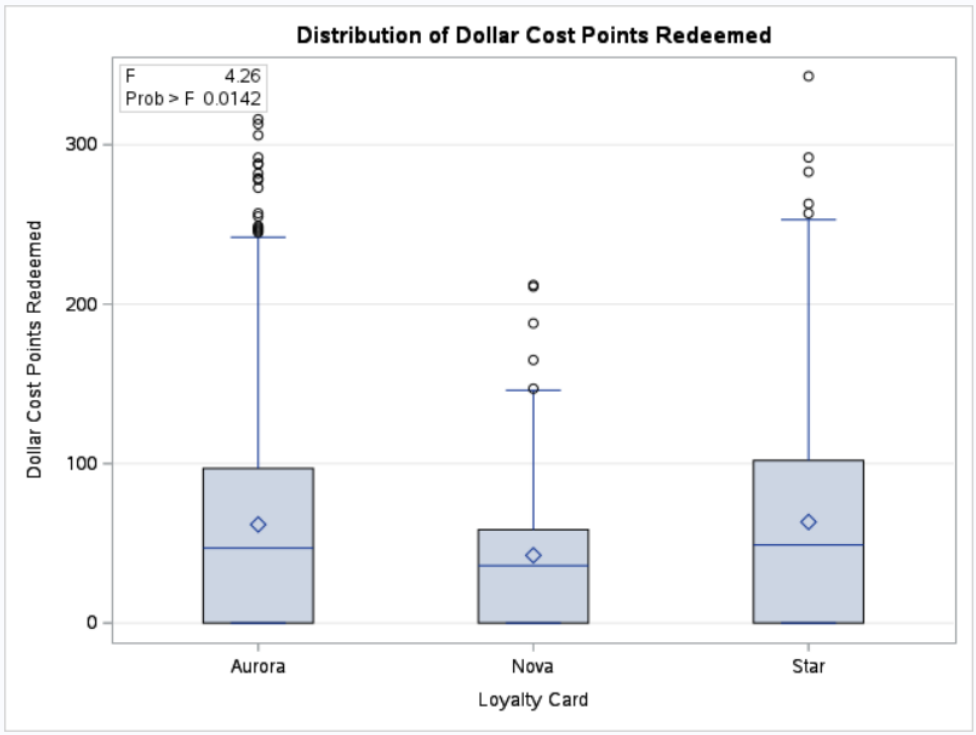


Figure 54: ANOVA Outputs for Differences in Dollar Cost Points Redeemed Between Loyalty Card Type

The ANOVA Procedure

Dependent Variable: Points Accumulated

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	75842880	37921440	13.79	<.0001
Error	2997	8243797059	2750683		
Corrected Total	2999	8319639939			

R-Square	Coeff Var	Root MSE	Points Accumulated Mean
0.009116	53.79212	1658.518	3083.199

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Loyalty Card	2	75842880.44	37921440.22	13.79	<.0001

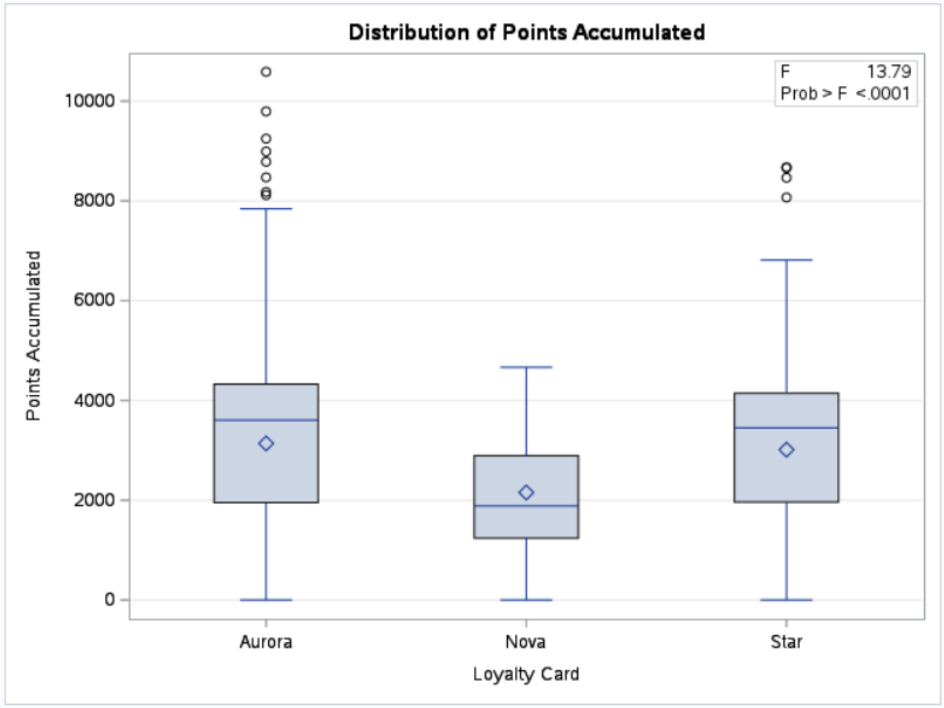


Figure 55: ANOVA Outputs for Differences in Points Accumulated Between Loyalty Card Type

The ANOVA Procedure

Dependent Variable: Total Flights

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	100287.50	50143.75	10.64	<.0001
Error	2997	14117601.77	4710.58		
Corrected Total	2999	14217889.27			

R-Square	Coeff Var	Root MSE	Total Flights Mean
0.007054	54.36702	68.63365	126.2413

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Loyalty Card	2	100287.5026	50143.7513	10.64	<.0001

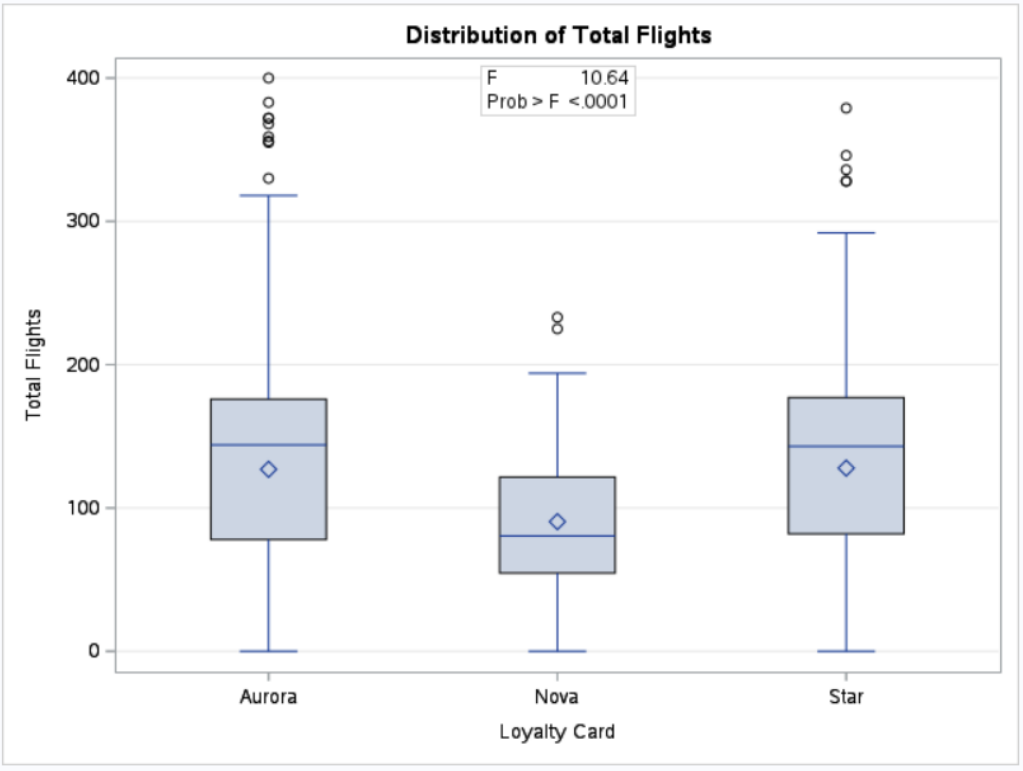


Figure 56: ANOVA Outputs for Differences in Total Flights Booked Between Loyalty Card Type

The ANOVA Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4935083008.6	2467541504.3	1.94	0.1444
Error	2236	2.8483453E12	1273857490.7		
Corrected Total	2238	2.8532804E12			

R-Square	Coeff Var	Root MSE	Salary Mean
0.001730	44.57554	35691.14	80068.89

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Loyalty Card	2	4935083009	2467541504	1.94	0.1444

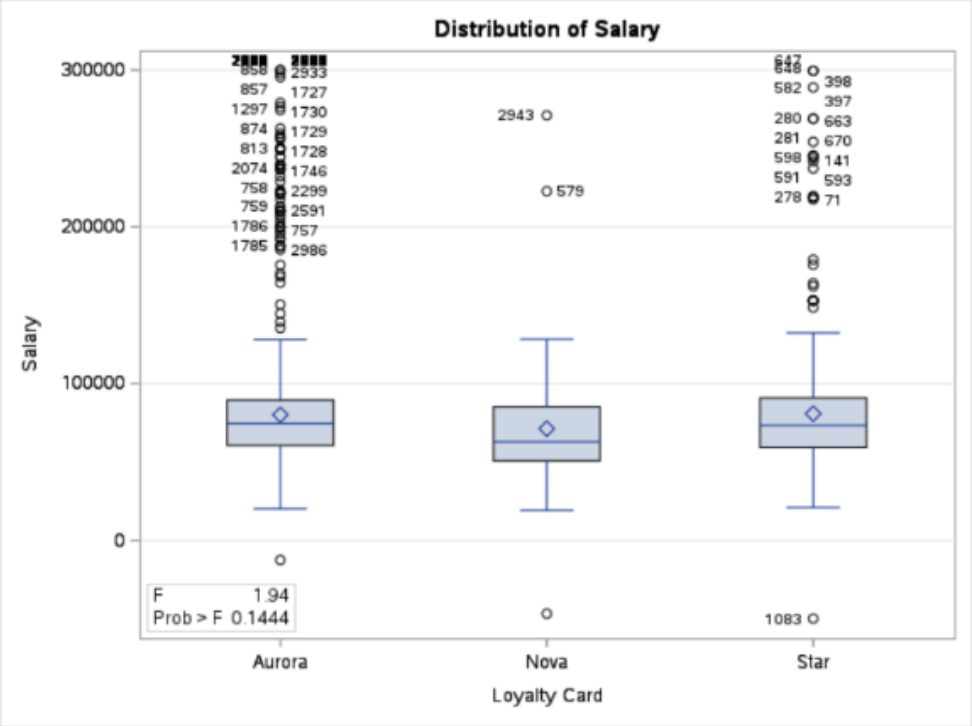


Figure 57: ANOVA Outputs for Differences in Salary Between Loyalty Card Type

The ANOVA Procedure

Dependent Variable: CLV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	12565099	6282549	1.48	0.2276
Error	2997	12713502048	4242076		
Corrected Total	2999	12726067147			

R-Square	Coeff Var	Root MSE	CLV Mean
0.000987	32.29119	2059.630	6378.303

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital Status	2	12565098.87	6282549.43	1.48	0.2276

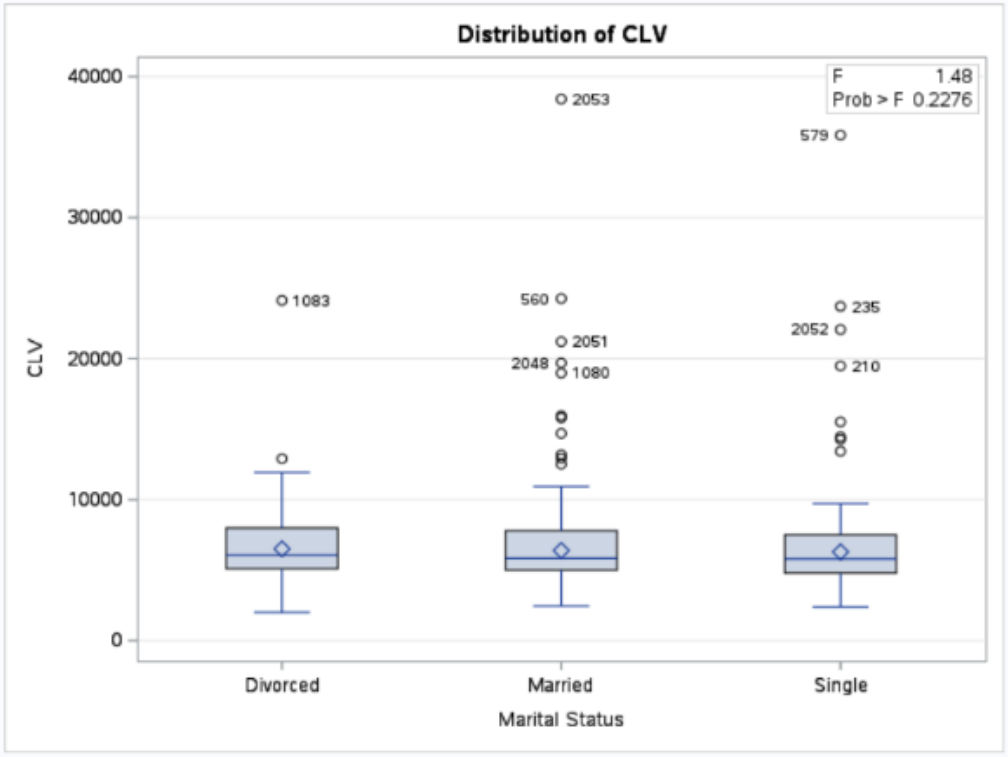


Figure 58: ANOVA Outputs for Differences in CLV Score Between Marital Statuses

The ANOVA Procedure

Dependent Variable: Dollar Cost Points Redeemed

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3297.65	1648.82	0.46	0.6292
Error	2997	10664698.71	3558.46		
Corrected Total	2999	10667996.36			

R-Square	Coeff Var	Root MSE	Dollar Cost Points Redeemed Mean
0.000309	96.68778	59.65281	61.69633

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital Status	2	3297.647220	1648.823610	0.46	0.6292

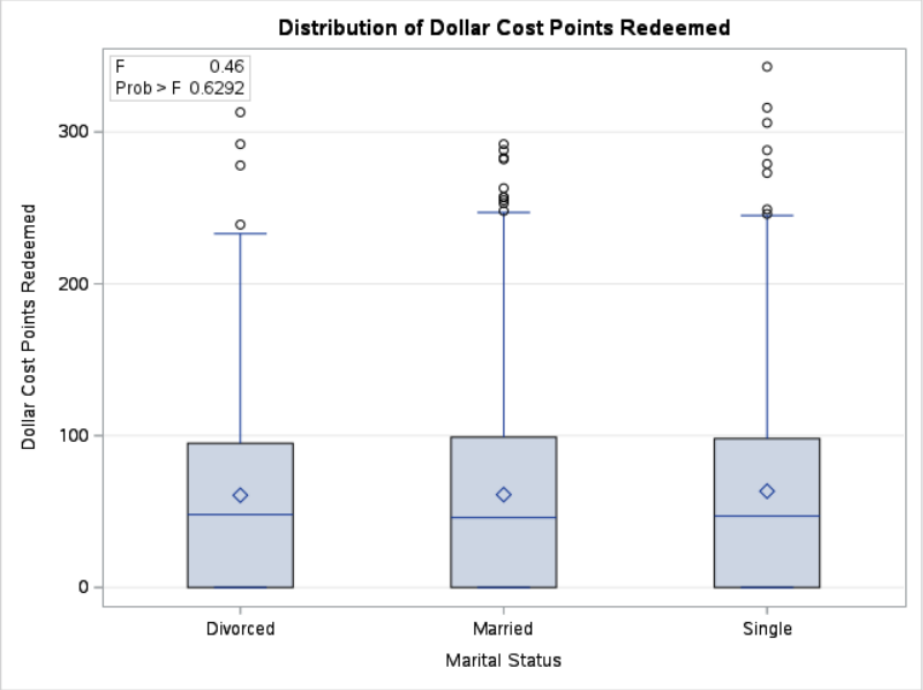


Figure 59: ANOVA Outputs for Differences in Dollar Cost Points Redeemed Between Marital Statuses

The ANOVA Procedure

Dependent Variable: Points Accumulated

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	10365048	5182524	1.87	0.1544
Error	2997	8309274892	2772531		
Corrected Total	2999	8319639939			

R-Square	Coeff Var	Root MSE	Points Accumulated Mean
0.001246	54.00532	1665.092	3083.199

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital Status	2	10365047.76	5182523.88	1.87	0.1544

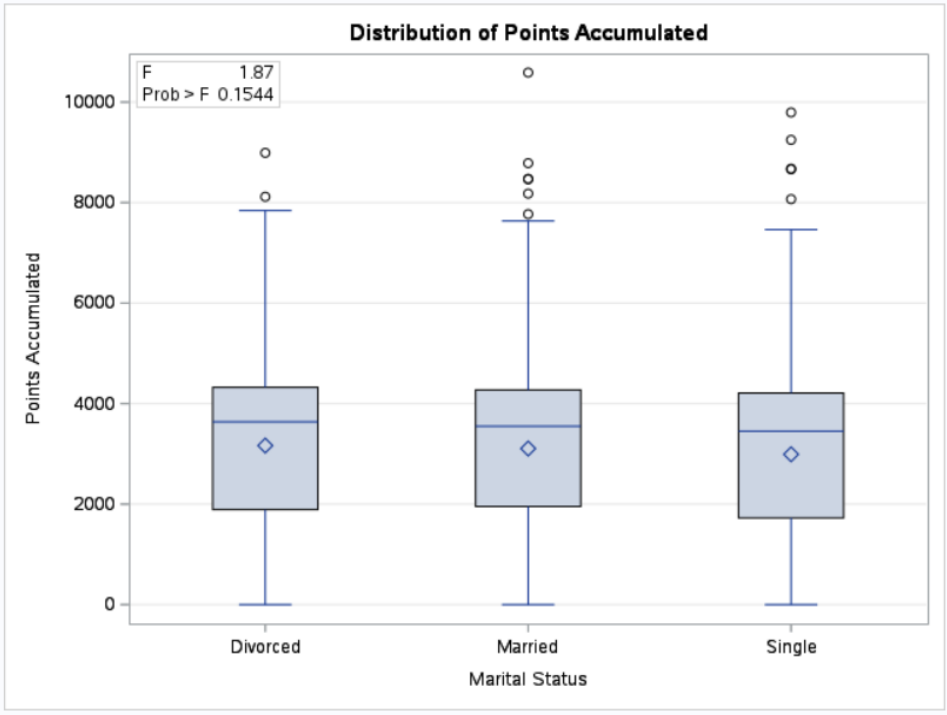


Figure 60: ANOVA Outputs for Differences in Points Accumulated Between Marital Statuses

The ANOVA Procedure

Dependent Variable: Total Flights

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2459.69	1229.85	0.26	0.7716
Error	2997	14215429.58	4743.22		
Corrected Total	2999	14217889.27			

R-Square	Coeff Var	Root MSE	Total Flights Mean
0.000173	54.55506	68.87104	126.2413

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital Status	2	2459.694716	1229.847358	0.26	0.7716

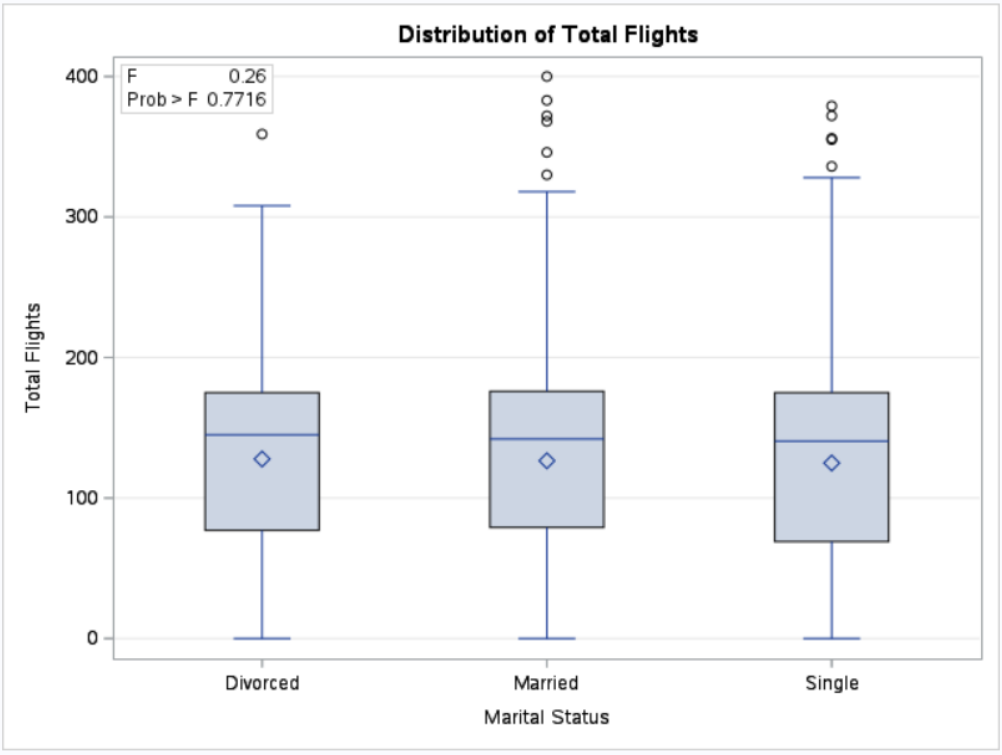


Figure 61: ANOVA Outputs for Differences in Total Flights Booked Between Marital Statuses

The ANOVA Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	27534216402	13767108201	10.89	<.0001
Error	2236	2.8257462E12	1263750543.7		
Corrected Total	2238	2.8532804E12			

R-Square	Coeff Var	Root MSE	Salary Mean
0.009650	44.39835	35549.27	80068.89

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Marital Status	2	27534216402	13767108201	10.89	<.0001

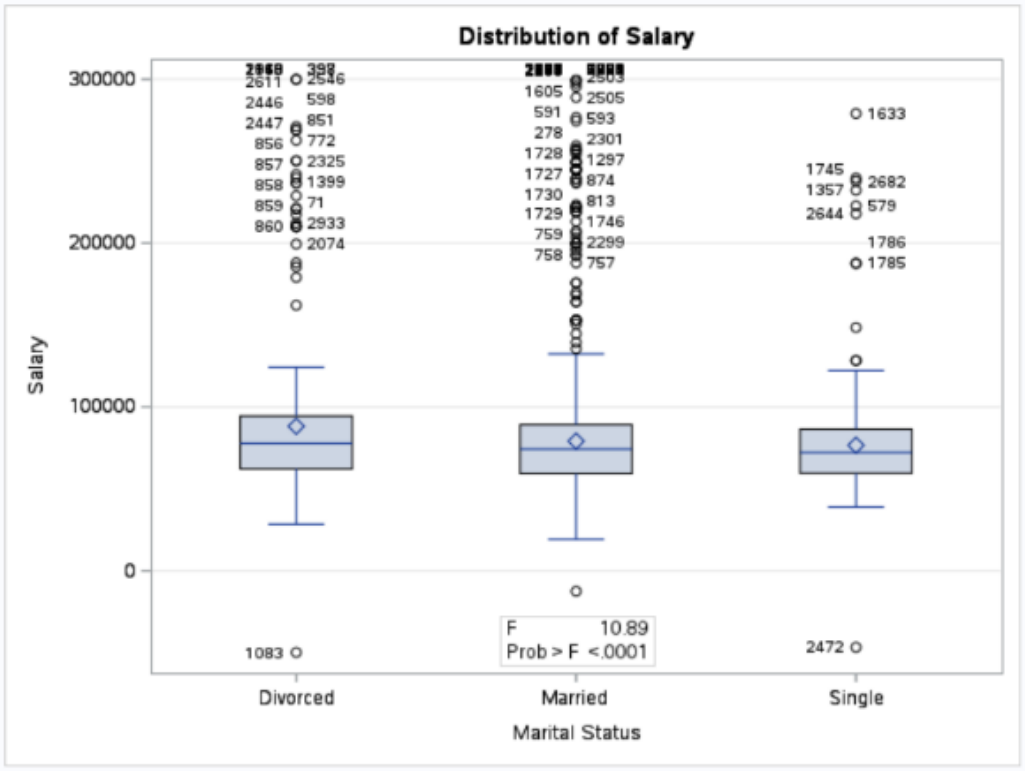


Figure 61: ANOVA Outputs for Differences in Salary Between Marital Statuses

3.3 Feature Engineering

One of the feature engineering techniques used was variable creation. As there were years and months for both enrollment and cancellation of airline loyalty programs by customers in the merged dataset, the cancellation status could be derived with two-factor levels being formed, 0 for not exited and 1 for exited. Referring to the figure below, this could be done under the condition that if the values for both cancellation year and month were missing, the cancellation status would be 0, or else it would be labelled as 1. After that, both cancellation year and month columns were dropped as the proportion of missing values in both columns

was as high as approximately 87%, in addition to the fact that imputing missing values, in this case, would be making no sense at all as this would assume that all enrolled customers had not exited the loyalty program. This could be used for further airline customer churn analysis.

	Cancellation Year	Cancellation Month	Cancellation Status
1	.	.	0
2	.	.	0
3	2018	1	1
4	.	.	0
5	.	.	0
6	.	.	0
7	.	.	0
8	.	.	0
9	.	.	0
10	.	.	0

Figure 62: First 10 rows for cancellation year, month and status

One hot encoding was also done on categorical variables such as gender, marital status, education level, enrolment type and loyalty card type, all of which were displayed in the figures below. To encode gender, two new binary variables were created based on the gender levels, male and female. If the gender value was male, the value for the Gender_Male column was set to 1 or else 0. If the gender value was female, the value for the Gender_Female column was set to 1 or else 0. To encode enrolment type, two new binary variables were formed based on the levels – Standard and 2018 Promotion. If the enrolment type value was Standard, the value for the Enrollment_Standard column was set to 1 or else 0. If the enrolment type value was 2018 Promotion, the value for the Enrollment_2018_Promotion column was set to 1 or else 0. For marital status, three new binary variables were formed based on the levels – married, single and divorced. If the marital status value was married, the value for the Marital_Married column was set to 1 or else 0. If the marital status value was single, the value for the Marital_Single column was set to 1 or else 0. If the marital status value was divorced, the value for the Marital_Divorced column was set to 1 or else 0. For loyalty card type, three new binary variables were formed based on the levels – Star, Aurora and Nova. If the loyalty card status was Star, the value for the Loyalty_Star column was set to 1 or else 0. If the loyalty card status was Aurora, the value for the Loyalty_Aurora column

was set to 1 or else 0. If the loyalty card status was Nova, the value for the Loyalty_Nova column was set to 1 or else 0. For education, five new binary variables were created – high school or below, college, bachelor, master and doctorate. If the educational level was high school or below, the value for Education_High_School_or_Below column was set to 1 or else 0. If the educational level was college, the value for the Education_College column was set to 1 or else 0. If the educational level was bachelor, the value for the Education_Bachelor column was set to 1 or else 0. If the educational level was master, the value for the Education_Master column was set to 1 or else 0. If the educational level was doctorate, the value for the Education_Doctorate column was set to 1 or else 0. The original variables, gender, marital status, education, enrolment type, and loyalty card, along with city and province were then dropped after the encoding process as city and province variables served a negligible purpose for the prediction task.

	Gender	Gender_Male	Gender_Female
1	Female	0	1
2	Male	1	0
3	Male	1	0
4	Male	1	0
5	Male	1	0
6	Male	1	0
7	Female	0	1
8	Male	1	0
9	Female	0	1
10	Male	1	0

Figure 63: First 10 rows for Gender, Gender_Male and Gender_Female

	Enrollment Type	Enrollment_2018_Promotion	Enrollment_Standard
1	Standard	0	1
2	Standard	0	1
3	Standard	0	1
4	Standard	0	1
5	Standard	0	1
6	Standard	0	1
7	Standard	0	1
8	Standard	0	1
9	Standard	0	1
10	Standard	0	1

Figure 63: First 10 rows for Enrollment Type, Enrollment_2018_Promotion and Enrollment_Standard

	Marital Status	Marital_Married	Marital_Single	Marital_Divorced
1	Married	1	0	0
2	Divorced	0	0	1
3	Single	0	1	0
4	Single	0	1	0
5	Married	1	0	0
6	Married	1	0	0
7	Single	0	1	0
8	Married	1	0	0
9	Married	1	0	0
10	Married	1	0	0

Figure 64: First 10 rows for Marital Status, Marital_Married, Marital_Single, and Marital_Divorced

	Loyalty Card	Loyalty_Star	Loyalty_Aurora	Loyalty_Nova
1	Star	1	0	0
2	Star	1	0	0
3	Star	1	0	0
4	Star	1	0	0
5	Star	1	0	0
6	Star	1	0	0
7	Star	1	0	0
8	Star	1	0	0
9	Star	1	0	0
10	Star	1	0	0

Figure 65: First 10 rows for Loyalty Card, Loyalty-Star, Loyalty_Aurora, and Loyalty_Nova

	Education	Education_High_School_or_Below	Education_College	Education_Bachelor	Education_Master	Education_Doctorate
1	Bachelor	0	0	1	0	0
2	College	0	1	0	0	0
3	College	0	1	0	0	0
4	College	0	1	0	0	0
5	Bachelor	0	0	1	0	0
6	Bachelor	0	0	1	0	0
7	College	0	1	0	0	0
8	Bachelor	0	0	1	0	0
9	Bachelor	0	0	1	0	0
10	Bachelor	0	0	1	0	0

Figure 66: First 10 rows for Education, Education_High_School_or_Below, Education_College, Education_Bachelor, Education_Master and Education_Doctorate

Outliers were identified in the latest merged dataset using boxplots and corresponding statistics using the PROC PRINT function and were then replaced with missing values for further imputation. For Flights Booked, any data greater than or equal to 265 were outliers. For Flights with Companions, any data greater than or equal to 71 were outliers. For Total Flights, any data greater than or equal to 328 were outliers. For Distance, any data greater than or equal to 78,159 were outliers. For Points Accumulated, any data greater than or equal to 8,068.68 were outliers. For Points Redeemed, any data greater than or equal to 3034 were outliers. For Dollar Cost Points Redeemed, any data greater than or equal to 246 were outliers. For Salary, any data greater than or equal to 135,464 and less than zero were outliers. For CLV, any data greater than or equal to 12,516.92 were outliers. However, for enrollment and cancellation year and month, all four columns had no outliers.

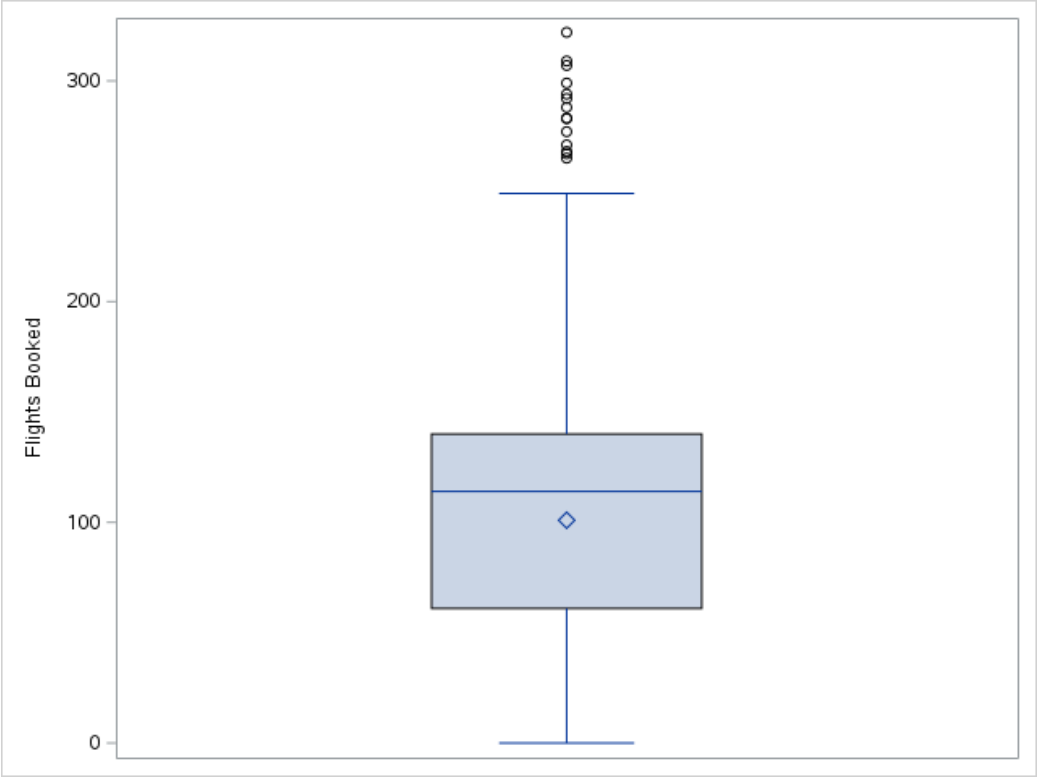


Figure 67: Boxplot for Flights Booked

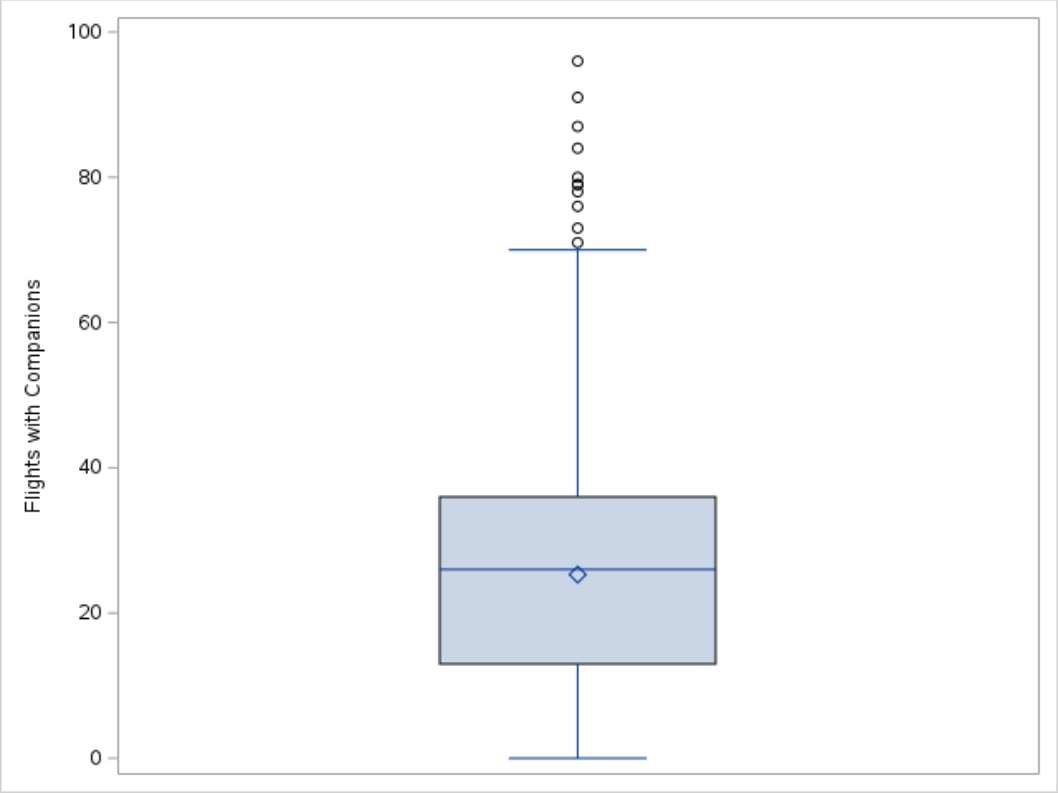


Figure 68: Boxplot for Flights with Companions

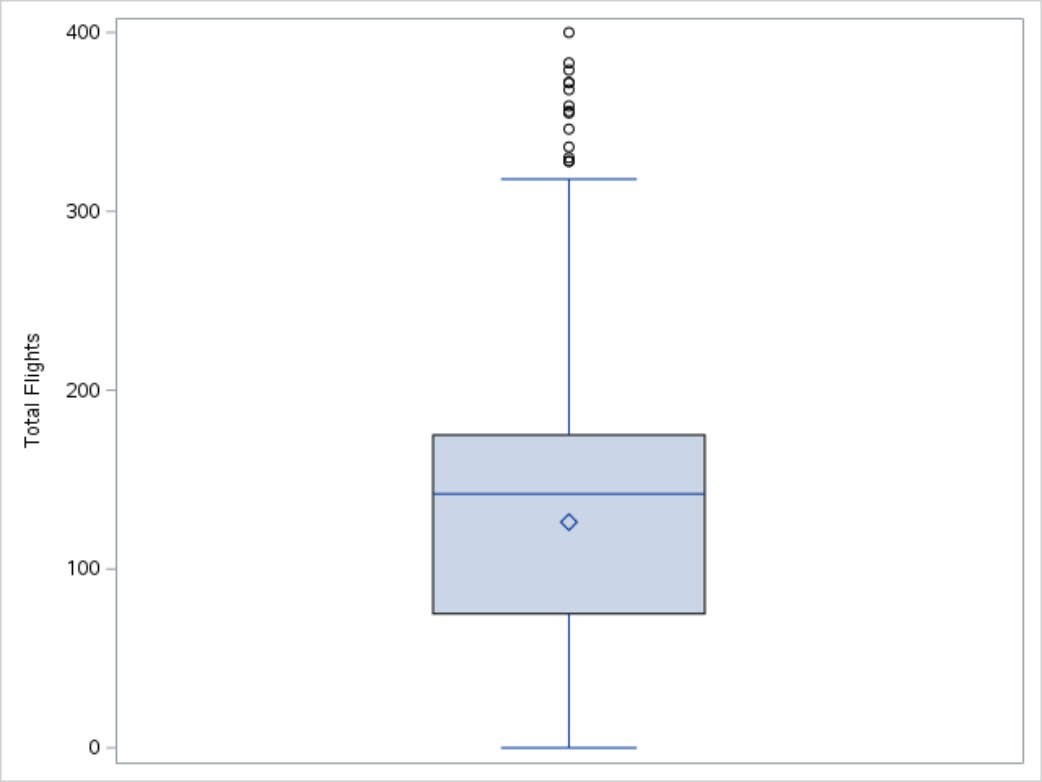


Figure 69: Boxplot for Total Flights

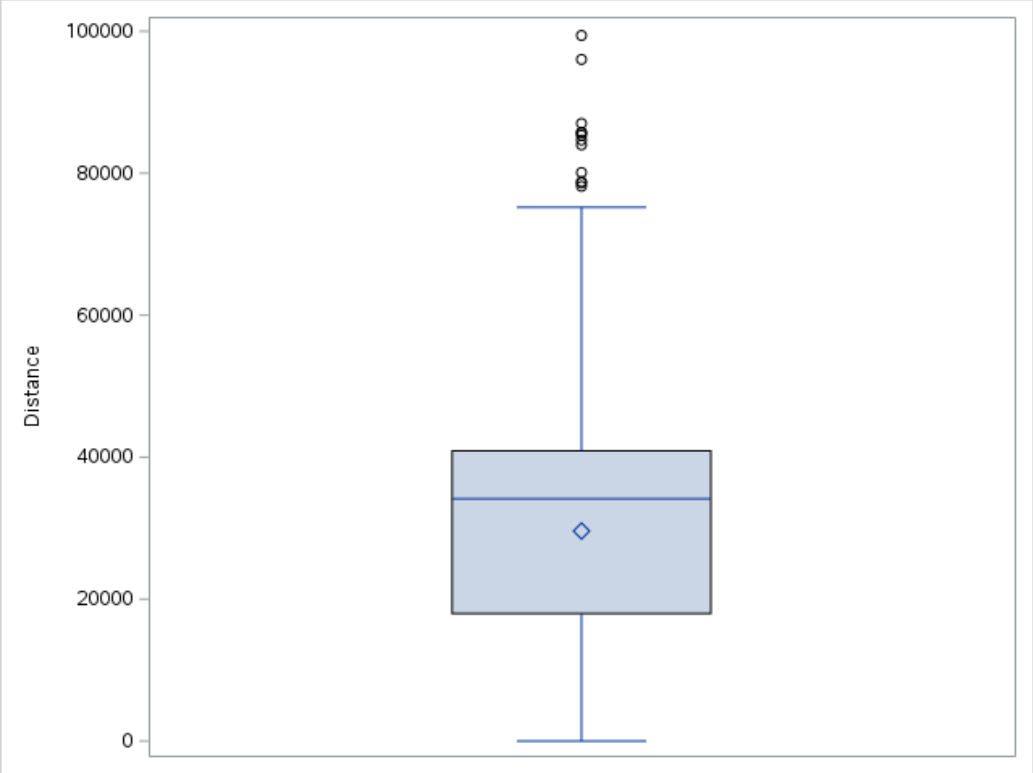


Figure 70: Boxplot for Distance

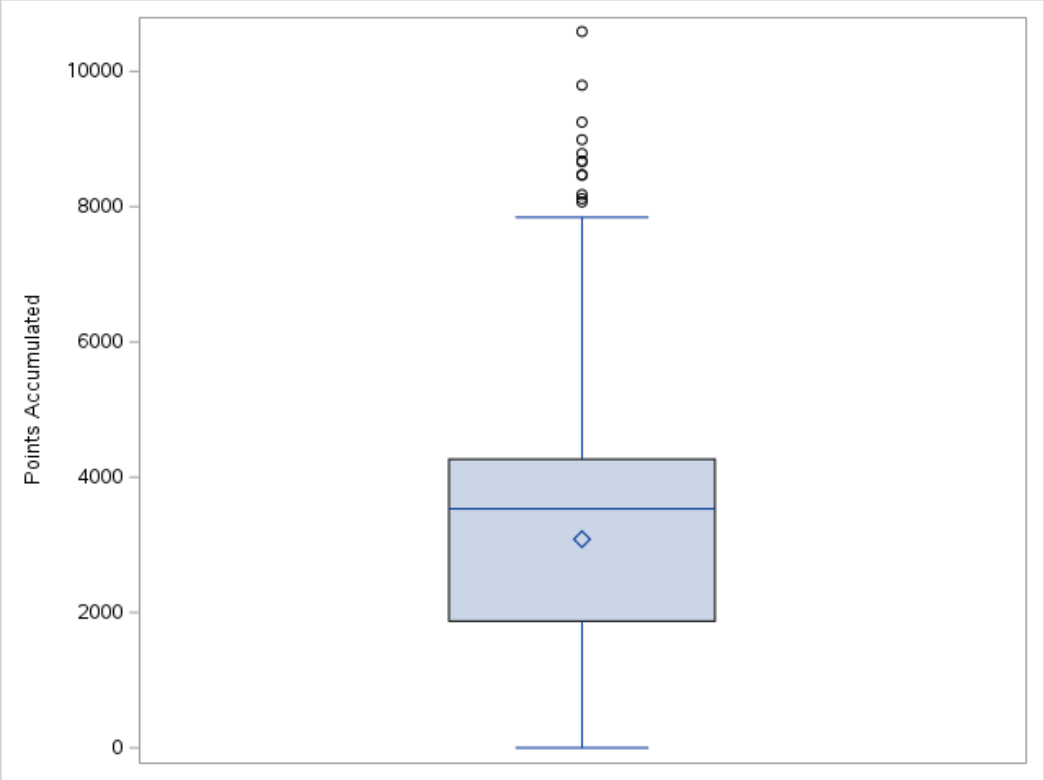


Figure 71: Boxplot for Points Accumulated

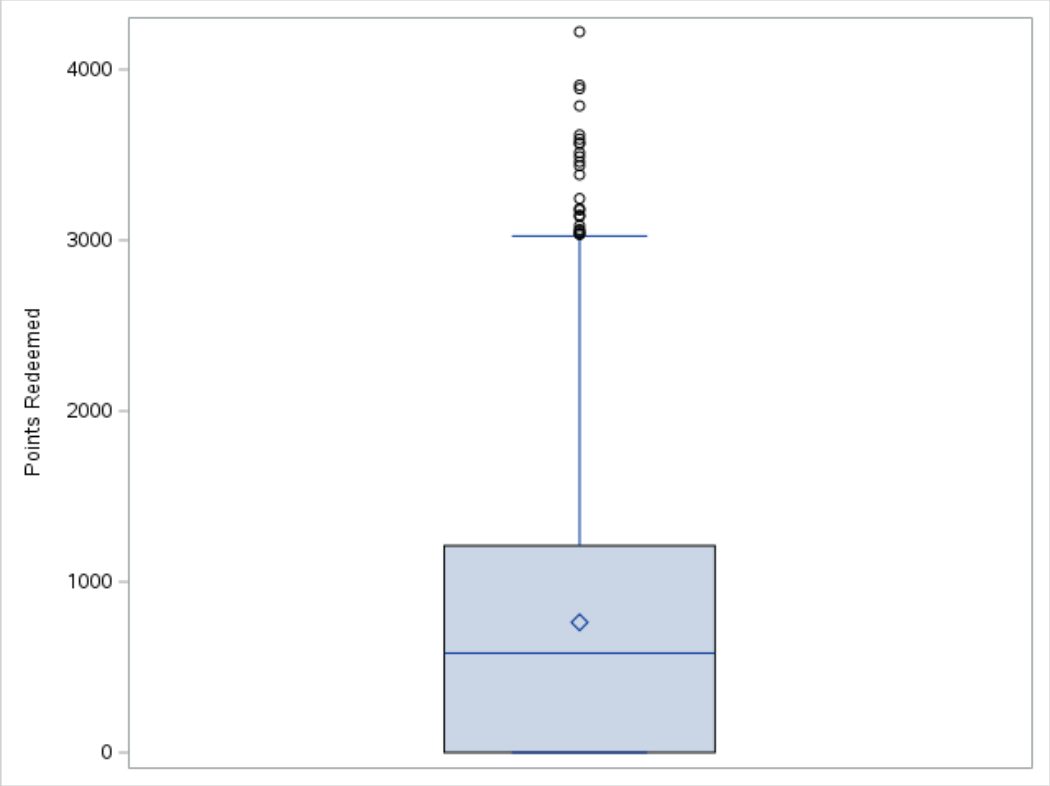


Figure 72: Boxplot for Points Redeemed

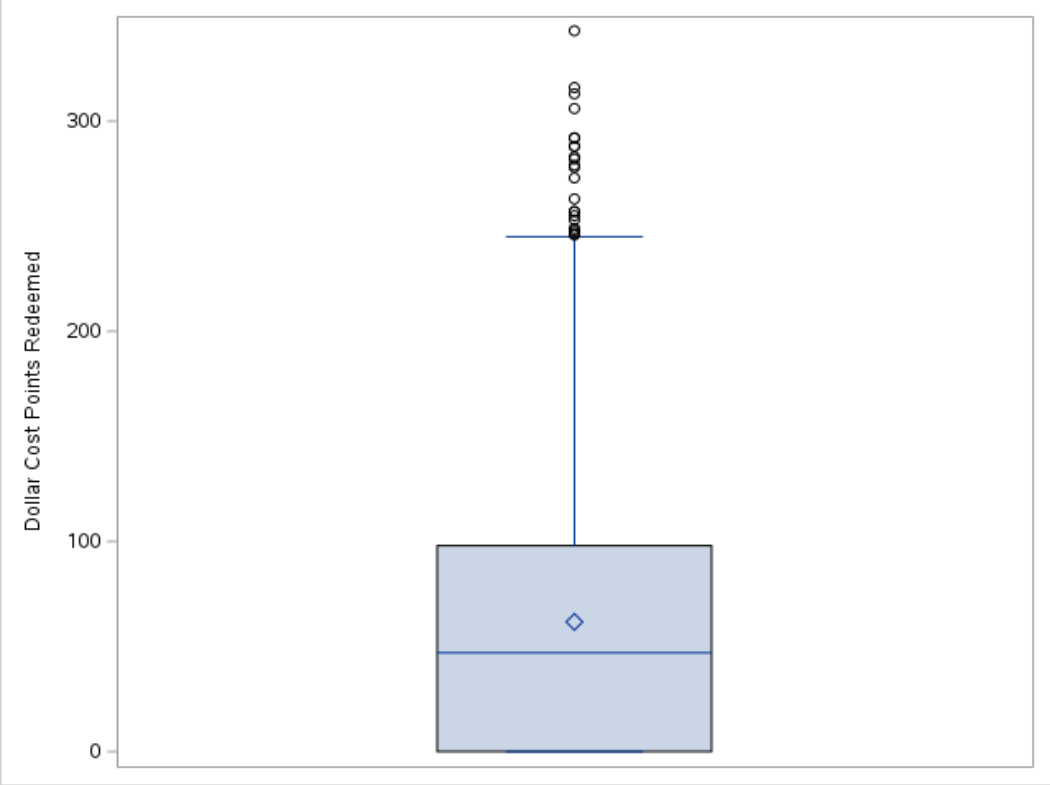


Figure 73: Boxplot for Dollar Cost Points Redeemed

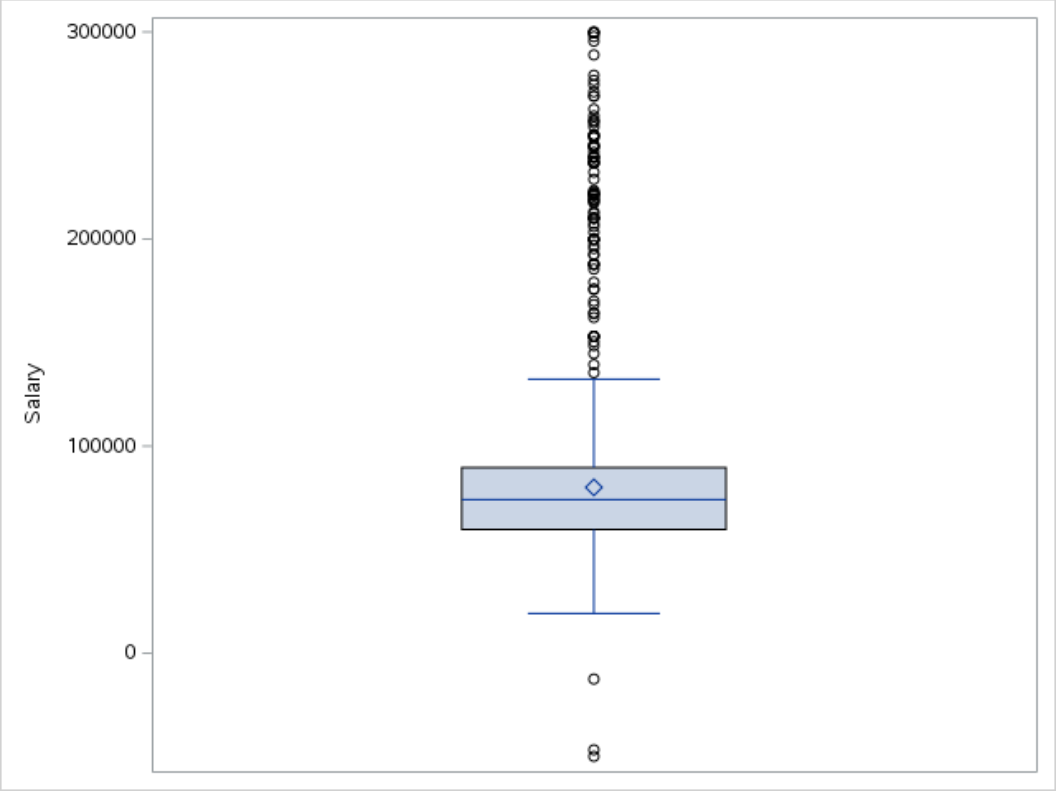


Figure 74: Boxplot for Salary

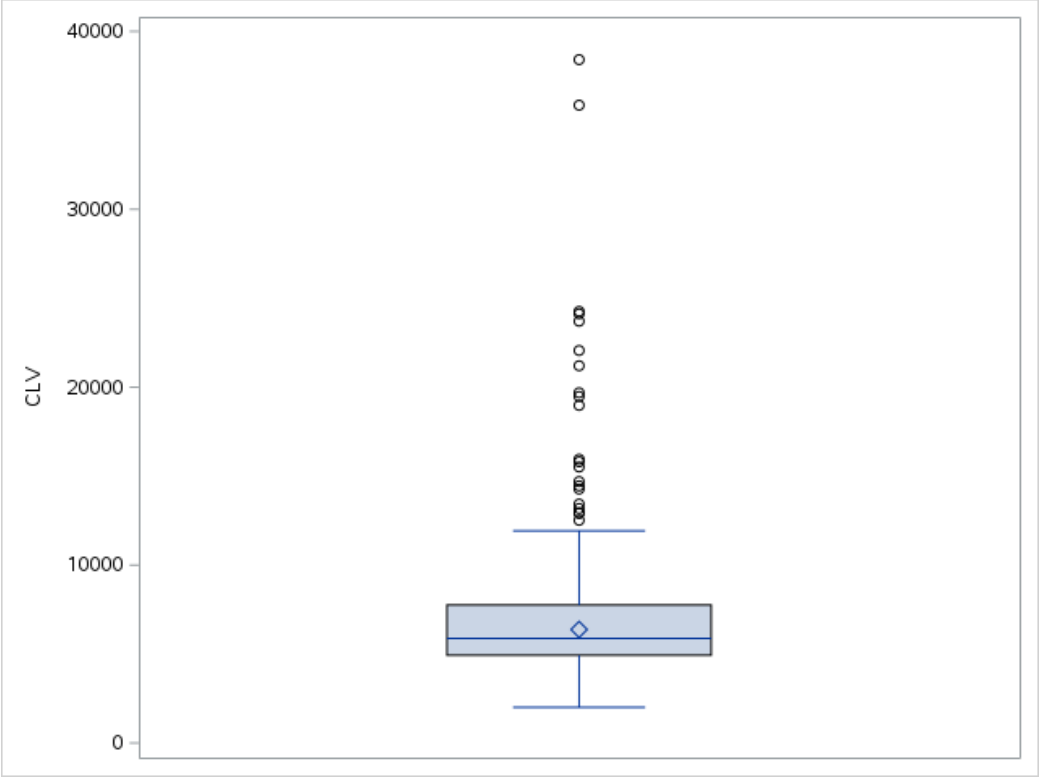


Figure 75: Boxplot for CLV

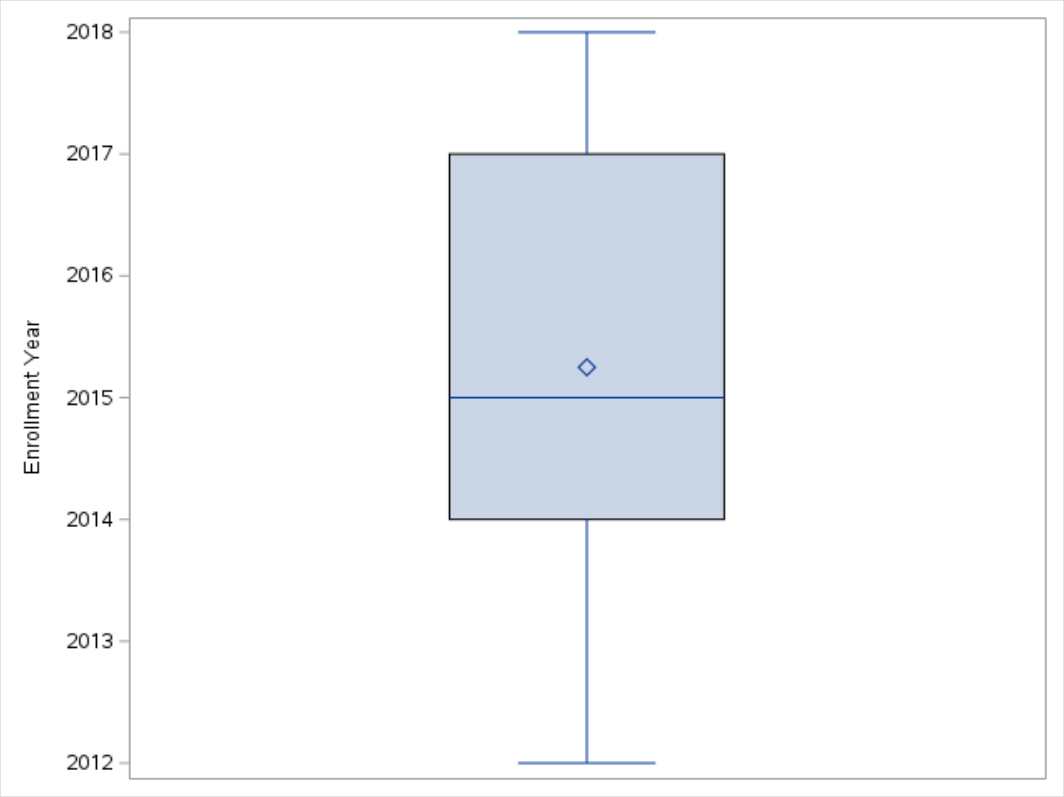


Figure 76: Boxplot for Enrollment Year

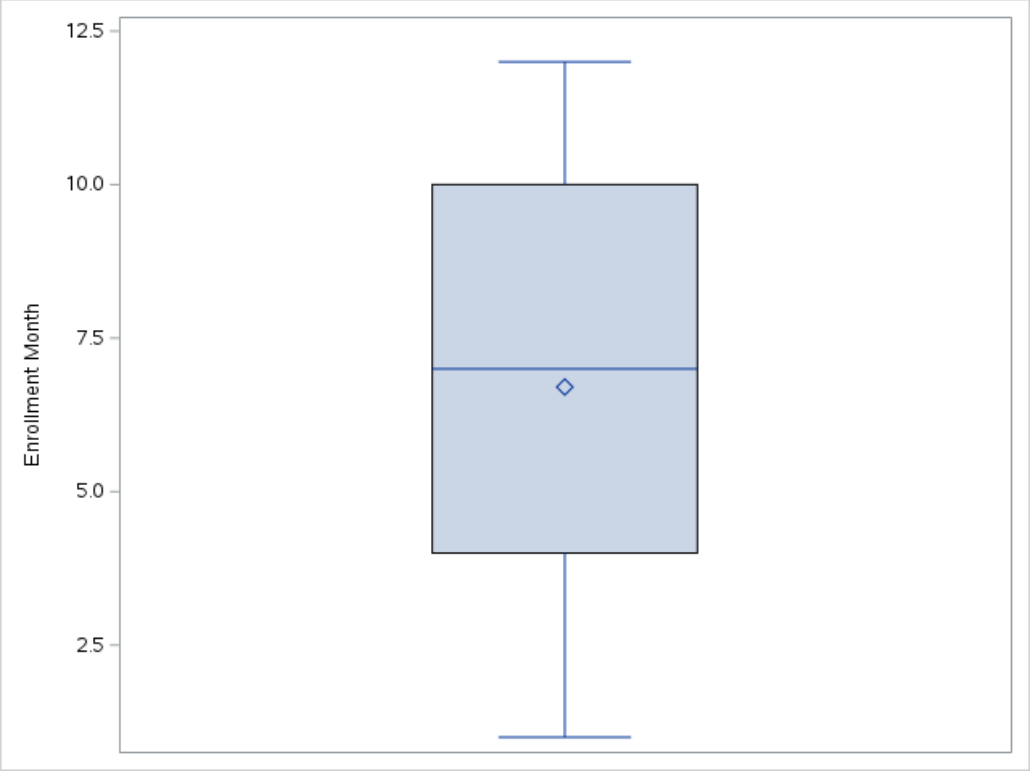


Figure 77: Boxplot for Enrollment Month

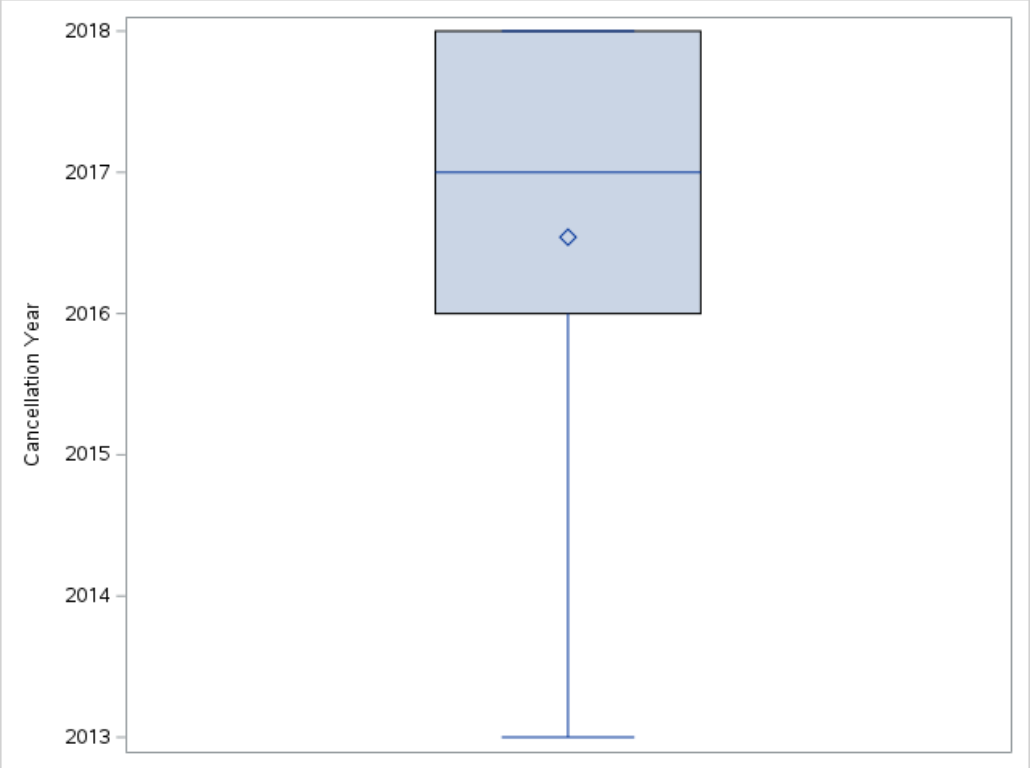


Figure 78: Boxplot for Cancellation Year

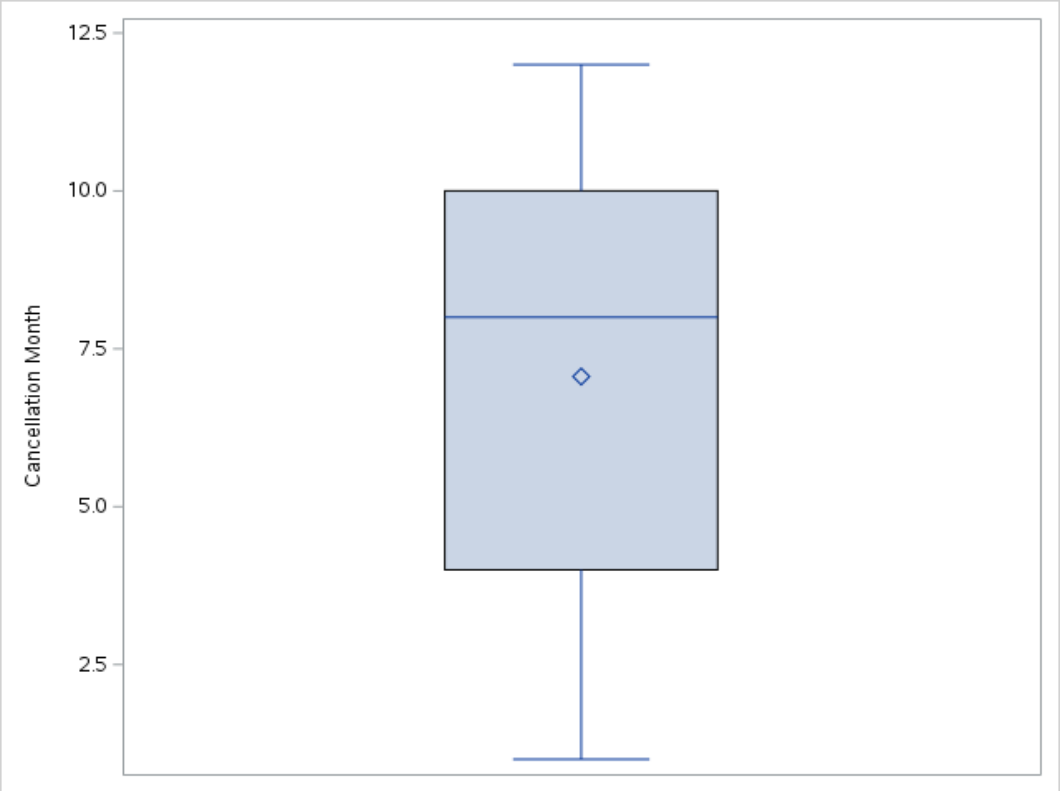


Figure 79: Boxplot for Cancellation Month

After handling the outliers, the missing values in each numerical variable needed to be imputed with median values as it was less sensitive to outliers compared to mean imputation.

The MEANS Procedure

Variable	N Miss
Flights Booked	14
Flights with Companions	11
Total Flights	14
Distance	12
Points Accumulated	11
Points Redeemed	24
Dollar Cost Points Redeemed	23
Salary	761
CLV	21
Enrollment Year	0
Enrollment Month	0
Gender_Male	0
Gender_Female	0
Enrollment_2018_Promotion	0
Enrollment_Standard	0
Marital_Married	0
Marital_Single	0
Marital_Divorced	0
Loyalty_Star	0
Loyalty_Aurora	0
Loyalty_Nova	0
Education_High_School_or_Below	0
Education_College	0
Education_Bachelor	0
Education_Master	0
Education_Doctorate	0

Figure 80: Number of missing values in the merged dataset after replacing outliers with missing values

Log transformation was also done as most of the histogram distributions of the numerical variables were right-skewed. Log (x + 1) was used instead of log(x) as there were zero values in each of the numerical variables. After the log transformation, nine new variables, better known as the log version of each numerical variable were created.

CLV_log	Distance_log	Dollar_Cost_Points_Redeemed_log	Flights_Booked_log	Flights_with_Companions_log	Points_Accumulated_log	Points_Redeemed_log	Salary_log	Total_Flights_log
8.25326	10.8567	4.75359	4.89035	3.68888	8.56129	7.25771	11.3294	5.14749
8.25339	10.6354	5.07517	5.25227	3.25810	8.33736	7.58680	11.2142	5.37528
8.25342	9.8866	3.43399	4.20469	3.09104	7.58274	5.92693	11.2142	4.47734
8.25342	10.4925	4.66344	4.82028	3.61092	8.19635	7.16395	11.2142	5.07517
8.25421	10.5144	0.00000	4.89035	3.80666	8.21357	0.00000	11.5473	5.17615
8.25468	10.4359	4.60517	4.74493	4.00733	8.16735	7.10906	10.8420	4.96284
8.25815	10.7850	4.85981	4.96981	3.25810	8.49327	7.36771	11.2142	5.12990
8.25907	10.6281	4.09434	4.97673	3.73767	8.33781	6.59851	11.5145	5.22575
8.25907	10.4596	4.82028	4.52179	3.52636	8.16906	7.32449	11.5145	4.82831
8.25907	10.4230	4.92725	5.01064	3.82864	8.13571	7.42893	11.5145	5.27300

Figure 81: First 10 rows for the log (x + 1) version of numerical variables

Finally, the data of numerical variables were normalized to the range between 0 and 1, and the merged dataset was saved as WORK.Flight_Loyalty_Normalized for easier reference.

Flights Booked	Flights with Companions	Total Flights	Distance	Points Accumulated	Points Redeemed	Dollar Cost Points Redeemed	Salary	CLV
0.53012	0.55714	0.53774	0.68979	0.64750	0.46876	0.46939	0.3804244346	0.1849139321
0.76305	0.35714	0.67610	0.55285	0.51756	0.65157	0.64898	0.3545140844	0.1849612997
0.26506	0.30000	0.27358	0.26147	0.24329	0.12364	0.12245	0.3545140844	0.1849754092
0.49398	0.51429	0.50000	0.47925	0.44948	0.42678	0.42857	0.3545140844	0.1849754092
0.53012	0.62857	0.55346	0.48985	0.45728	0.00000	0.00000	0.4383432014	0.1852817867
0.45783	0.77143	0.44654	0.45286	0.43663	0.40397	0.40408	0.2886189437	0.1854611787
0.57430	0.35714	0.52830	0.64212	0.60491	0.52331	0.52245	0.3545140844	0.1868096428
0.57831	0.58571	0.58176	0.54886	0.51779	0.24231	0.24082	0.4288058596	0.1871664114
0.36546	0.47143	0.38994	0.46376	0.43738	0.50116	0.50204	0.4288058596	0.1871664114
0.59839	0.64286	0.61006	0.44706	0.42302	0.55636	0.55918	0.4288058596	0.1871664114

Figure 82: First 10 rows of normalized variables

4.0 Hypothesis

The following five hypotheses are formulated based on the analysis of the cleaned dataset:

1. There is a significant association between the type of loyalty program enrolment and CLV.
2. There is a significant association between the status of loyalty and CLV.
3. There is no association between marital status and total number of flight reservations.
4. The total number of flight reservations does not significantly predict CLV.
5. There is no relationship between the salary of customers and the total number of flight reservations.

5.0 Discussion and Conclusion

To summarise, there are statistically significant relationships between loyalty card status and enrolment type. This can indicate that if flyers subscribe to the loyalty program through standard means or special promotional occasions, they are more likely to do their best to climb up the ladder of loyalty status so that they can enjoy exclusive incentives from the airline without spending much on the airline services like flight ticket and discounted usage of premium lounge services. Although CLV negligibly correlates with other variables, it is

indicated to be significantly associated with enrolment type and loyalty card status based on a chi-square test and two-sample t-test, suggesting that flyers enrolled in the loyalty program possessing higher loyalty status have high CLV and share more trust with the airline in terms of quality-of-service delivery compared to those not in the program. However, gender does not play a significant role in predicting other variables based on the analysis, suggesting that male and female passengers equally enjoy the benefits of the loyalty program. The same goes for salaries as customers may not prefer travelling and prefer shifting their financial resources towards other non-travel things.

Surprisingly, there is no difference in total flight bookings between martial statuses as travelling is one of the ways to enjoy leisure with friends, close ones and even themselves alone. The negligible correlation between the total number of flight bookings and the CLV of flyers is also surprising as flight reservations are typically one of the main sales channels for an airline company, and based on this as a metric, the company can focus on those who make frequent bookings that generate the most value. Therefore, it will be interesting to further examine the relationship between the total number of flight bookings and the other two variables, CLV and marital status.

To conclude, one of the aspects I gained in this assignment is using SAS to perform data preprocessing and EDA as other programming tools like Python and R are preferred by the data science community, so this assignment can help other readers to explore SAS more and adding it to their data science inventory. Another aspect is learning how to deal with outliers as one of the easiest ways to handle it is detecting or deleting them. Instead, I can treat them as missing values and impute them with median values, so readers can learn how to use SAS to deal with outliers.

References

Batarlienė, N., & Slavinskaitė, N. (2023). Assessment of factors determining airline consumer loyalty: Case study in Lithuania. *Sustainability*, *15*(2), 1320. https://doi.org/10.3390/su15021320

Bravo, A., & Vieira, D. R. (2019). A systematic review of the civilian airline industry: towards a general model of customer loyalty. *International Journal of Business and Data Analytics*, *1*(2), 156-183. https://doi.org/10.1504/IJBDA.2019.104162

de Jong, G., Behrens, C., & van Ommeren, J. (2019). Airline loyalty (programs) across borders: A geographic discontinuity approach. *International Journal of Industrial Organization*, *62*, 251-272. https://doi.org/10.1016/j.ijindorg.2018.02.005

Gao, Y., Carrigg, M., Lewinski, R., Polderman, D., & Tkalcevic, P. (2018). The perceived value of frequent flyer program benefits among Australian travelers. *International Journal of Aviation, Aeronautics, and Aerospace*, *5*(3), 6. https://doi.org/10.15394/ijaaa.2018.1249

Koech, A. K., Buyle, S., & Macário, R. (2023). Airline brand awareness and perceived quality effect on the attitudes towards frequent-flyer programs and airline brand choice-Moderating effect of frequent-flyer programs. *Journal of Air Transport Management*, *107*, 102342. https://doi.org/10.1016/j.jairtraman.2022.102342

Lim, F., Chun, S. Y., & Satopää, V. (2023). Loyalty currency and mental accounting: Do consumers treat points like money? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3974642

Limberger, P. F., Pereira, L. A., & Pereira, T. (2021). The impact of customer involvement in airline loyalty programs: a multigroup analysis. *Tourism & Management Studies*, *17*(3), 37-49. https://doi.org/10.18089/tms.2021.170303

Orhun, A. Y., Guo, T., & Hagemann, A. (2022). Reaching for gold: Frequent-flyer status incentives and moral hazard. *Marketing Science*, *41*(3), 548-574. https://doi.org/10.1287/mksc.2021.1341

Ruibin, K. D., & Borglöv, T. V. (2018). Predicting customer lifetime value: understanding its accuracy and drivers from a frequent flyer program perspective. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1234459

Thirumuruganathan, S., Jung, S. G., Ramirez Robillos, D., Salminen, J., & Jansen, B. J. (2021). Forecasting the nearly unforecastable: why aren't airline bookings adhering to

the prediction algorithm?. *Electronic Commerce Research*, *21*, 73-100. https://doi.org/10.1007/s10660-021-09457-0