Summer 2014

Bryan Epstein
blespooky@hotmail.com
052-5249926

Exploring the Gap: operator provided quality of service compared to that which is desired by the passenger in a contract tendering regime

Passenger Demand Summary: results of *Transportation Today and Tomorrow's* full survey.

Analysis objective: ascertain through a battery of statistical and analytical quantitative methods which quality of service (QOS) parameters are most important for public transportation (PT) users. This knowledge is important for two reasons, first in its ability to help policy best serve the PT using public and second, by providing a body of knowledge with which to compare to the current and future QOS supply as defined by the current contract tendering regime in Israel.

The following table lists the following subjects of analysis, page numbers and what parameters where most important for the subject.

Subject	Page	Important Parameters	Notes
I. *Socio Demographic Summary*	1	-	The survey appears to have reached a normally distributed example of PT user's.
II. *Ride Characteristics Summary*	2	-	Rider characteristics are as expected. There is growth well above natural growth in bus use and people believe that without PT they cannot complete their journey.
III. *Complaint Handling*	3	-	Very few PT users participated such that the impact of complaint analysis is limited.
IV. *Data Validation*	3	-	The 14 quality parameters are presumed to express different characteristics of overall satisfaction according to statistical tests.
V. *Data Transformation*	4	-	Data transformation unfortunately does not improve the distribution of the 14 quality parameters; the non-transformed data will be used.

Subject	Page	Important Parameters	Notes
VI. *Central Station Data*	5	-	Central region accounts for 48% of data offering different options for mapping and interpolation of QOS.
VII. *Compare Means*	6	Frequency, Comfort, Fare, DirectRte	Lack of variance in overall satisfaction shows that the survey population's responses are congruent.
VIII. *Importance and Satisfaction*	7	RideTime, Fare, OnTime, Crowded, Frequency	When graphed, objective parameters are more important with lower than average satisfaction scores.
IX. *Improvement and Satisfaction*	8	RideTime, Fare, OnTime, Frequency, Crowded	The impact of improvements of quality parameters are perceived as higher for parameters with lower average satisfaction.
X. *Importance, Improvement and Satisfaction*	9	Frequency, OnTime, RideTime, Fare, DirectRte, Crowded	It appears that as overall satisfaction for a parameter improves, its perceived importance and improvement ability decreases.
XI. *Linear Regression Models for Predicting Overall Satisfaction*	10	Frequency, OnTime, Transfer, Driving, RideTime	Objective factors have the strongest ability to predict overall satisfaction in a single variable regression model.
XII. *Multi Variable Regression*	11	Frequency, RideTime, Comfort, Driving, OnTime, Distance	Six parameters were entered achieving an Rsquare value of .516 which is moderately strong.
XIII. *Principle Component Analysis*	12	-	Two components explain approximately 48% of the data, additional components have little effect on variance explained.
XIV. *Principle Component Analysis Continued*	13	Frequency, Transfer, Comfort, RideTime, Driving, DirectRte, OnTime	As PCA1 values increase, overall satisfaction grows, this shows that the variables included in this component have the ability to group respondents by their level of overall satisfaction.
XV. *Multi Variable Regression based on PCA1/2*	14	Frequency, Transfer, Comfort, RideTime, Driving, DirectRte, OnTime	-

Subject	Page	Important Parameters	Notes
XVI. *Multi Variable Regression based on PCA1/2 Continued*	15	-	This model shows how the PCA selected parameters in groups, one more objective and the other subjective.

Important Parameter Frequency in Statistical Tests

	Subject								
	VII	VIII	IX	X	XI	XII	XIV	XV	Sum
Safety									0
Distance						1			1
Frequency	1	1	1	1	1	1	1	1	8
Transfer					1		1	1	3
DriverBhv									0
Comfort	1					1	1	1	4
Fare	1	1	1	1					4
Ticket									0
RideTime		1	1	1	1	1	1	1	7
Driving					1	1	1	1	4
DirectRte	1			1			1	1	4
Crowded		1	1	1					3
OnTime		1	1	1	1	1	1	1	7
Info									0

- Frequency, ride time and reliability (OnTime) are by far the parameters which were shown by the different tests to be the most important.
- There appear to be two groups of parameters, they can be labeled objective and subjective for these purposes. The objective parameters are quantitatively measurable and appear to be more important, have the ability to improve most and are regularly rated as the least satisfying aspects of PT user's journey.
- Based on this survey of a test population, objective parameters should be the focus of supply rationing and improvements.

I. *Socio Demographic Summary*

-Gender – it is accepted that more women ride public transportation than men, this is certainly the case for this survey.

-Age Group – normal distribution

-Education – normal distribution

-Driver License – half of the surveyed population has a drivers license.

-Cars Owned – around 27% of respondents do not own a car while more than half own one car.

-Car Available – 61% of respondents do not have an available vehicle.

-House Population – surprisingly high frequency of respondents at the highest values of the survey (5-7+). Most respondents are living with one other person.

-Religion – the survey did not reach a significant Arab population.

-Religious Status – most respondents are secular, this may be explained by the survey being distributed in certain locations which are more 'secular'.

-Employment Status – most respondents are employees or students.

-Average income is mainly below average or average, above average bus riders are rare.

Sub Conclusion:

The socio demographic data shows that the survey managed to reach a normal distribution of respondents or at least percentages which are expected when surveying PT users. Most respondents are women, own at least one car, are using PT to get to work or study are Jewish, secular and have an average or below average income.

Socio Demographic Summary

		N	Percent
Gender	Male	212	36.4
	Female	370	63.6
	Total	582	
Age_Grp	15-18	56	9.4
	24-19	174	29.1
	44-25	192	32.2
	64-45	104	17.4
	65+	71	11.9
	Total	597	
Education	Partial	61	11.2
	High School	220	40.3
	Post High School	103	18.9
	Academic	162	29.7
	Total	546	
Drv_License	Yes	307	51.4
	No	290	48.6
	Total	597	
Cars_Owned	0	138	26.5
	1	292	56
	2	60	11.5
	3	19	3.6
	4	7	1.3
	6	5	1
	Total	521	
Car_Available	Yes	196	31.9
	No	375	61
	Total	571	
House_Pop	1	43	7.8
	2	122	22
	3	87	15.7
	4	106	19.1
	5	103	18.6
	6	44	7.9
	7-12	49	8.800
	Total	554	
Religion	Jewish	497	88.6
	Arab	64	11.4
	Total	561	
Relig_Status	Secular	301	51.8
	Traditional	130	22.4
	Orthodox	103	17.7
	Charedi	47	8.1
	Total	581	
Empl_Status	Wage/Employee	271	47.1
	Independent	18	3.1
	Student	125	21.7
	Retired	75	13
	Homemaker	11	1.9
	Unemployed	30	5.2
	Soldier	45	7.8
	Total	575	
Avg_Income	Below Average	222	43.4
	Average	221	43.2
	Above Average	69	13.5
	Total	512	

II. *Ride Characteristics Summary*

-Region - Nearly half of the respondents are in the Central region.

-Hour - Most respondents were surveyed during peak hours.

-Trip Type - 57% of trips are inside the city.

-Trip Frequency - 71% of PT riders ride each day, very few utilize PT on a rare basis.

-Trip Frequency Group - 71% of the riders surveyed are considered regular PT users.

-Trip Frequency on Line - 61% of the users surveyed were riding a bus line they ride everyday.

-Trip Objective - 35% of those surveyed use the bus to get to work, in addition 20% are running errands and 20% are heading home.

-Frequency Change - 62% of respondents report no change in their riding behavior while almost 30% ride more.

-Ride No PT - 70% could not have completed their trip without PT.

-Information Satisfaction - nearly half are extremely satisfied with information availability.

PT on Shabbat - 43.6 respondents would absolutely ride PT on Shabbat, while 31% would absolutely not.

Sub Conclusion: Rider characteristics are as expected, there is growth well above natural growth in bus use and many people believe that without PT they cannot

Ride Characteristics			
		N	Percent
Region	North	142	23.1
	Center	298	48.5
	Jerusalem	92	15
	South	83	13.5
	Total	615	
Hour	6-7	49	8.2
	8	49	8.2
	9	49	8.2
	10	66	11
	11	38	6.3
	12	30	5
	13	34	5.7
	14	58	9.7
	15	38	6.3
	16	77	12.8
	17	75	12.5
	18-20	38	6.3
	Total	601	
Trip_Type	Intracity	308	56.9
	Intercity	233	43.1
	Total	541	
Trip_Freq	Everyday	435	70.7
	Once a Week	85	13.8
	2-3 a Month	33	5.4
	Once a Month	4	0.7
	Less than Monthly	29	4.7
	Rarely	29	4.7
	Total	615	
Trip_Freq_Grp	Regular	435	70.7
	Irregular	180	29.3
	Total	615	
Trip_Freq_Line	1	296	61.2
	2	111	22.9
	3-4	44	9
	5-6	33	5.4
	Total	484	
Trip_Obj	Work	213	34.6
	Army	37	6
	School	82	13.3
	Errands	122	19.8
	Shopping	12	2
	Entertainment	12	2
	Home	137	22.3
	Total	615	
Freq_Chng	No Prev Ride	37	6
	Yes_More	176	28.7
	Yes_Less	18	2.9
	No Change	382	62.3
	Total	613	
RideNoPT	Yes	155	30.3
	No	357	69.7
	Total	512	
Info_Sat	1-4	9	2.2
	5	13	3.3
	6	11	2.8
	7	46	11.5
	8	84	21
	9	43	10.8
	10	194	48.5
	Total	400	
PT_Shabbat	Absolutely	255	43.6
	Most Likely	75	12.8
	Maybe	46	7.9
	Not Likely	29	5
	Absolutely Not	180	30.8
	Total	585	

III. *Complaint Handling*

-Complaint - only 30.6% claim to know where and how to submit a complaint

-Complaint Submitted - only 7.5% have submitted a complaint.

-Complaint Handling Satisfaction - obviously there are those who had good and bad experiences with complaint handling, the data is inconclusive.

Sub Conclusion: While complaints are a good way for operators and the authority to receive direct customer feedback, very few PT users participate such that the impact is minimal.

Complaint Handling			
		N	**Percent**
Cmpl	Yes	180	30.6
	No	409	69.4
	Total	589	
CmplSubmit	No	533	92.5
	Yes	43	7.5
	Total	576	
Cmpl_Sat	1	17	33.3
	2	5	9.8
	4	1	2
	5	2	3.9
	6	4	7.8
	7	6	11.8
	8	6	11.8
	9	2	3.9
	10	8	15.7
	Total	51	

IV. *Data Validation*

Quality Parameter Validation					
		Before Transformation	**After Tranformation**	**Before Transformation**	**After Tranformation**
	Cronbach's Alpha	0.905	0.917	0.884	0.895
Mean Satisfaction	Overall	7.6	0.4539	—	—
	Safety	8.37	0.3063	8.19	0.3349
	Dist	8.09	0.3582	7.78	0.4084
	Freq	7.38	0.4602	7.12	0.4988
	Transfer	7.68	0.4052	7.74	0.4085
	Driver	8.11	0.3577	7.97	0.3818
	Comfort	7.86	0.4073	7.63	0.4391
	Fare	7.51	0.4302	7.21	0.4804
	Ticket	8.49	0.279	8.66	0.2589
	RideTime	7.31	0.4716	7.18	0.4944
	Driving	7.69	0.417	7.68	0.425
	DirectRte	7.88	0.3861	7.98	0.3769
	Crowded	7.04	0.52	6.6	0.5711
	OnTime	7.36	0.4663	7.11	0.5041
	Info	7.64	0.3986	7.64	0.4198
	N	**279**	**279**	**450**	**450**

Cronbach's Alpha tests internal consistency, or how well a group of measures test a similar construct. If the Cronbach's Alpha value is greater than 0.7 then the data is theoretically describing similar subjects. Cronbach's Alpha is best used when different measures express characteristics of a single subject. In this case the 14 quality parameters are presumed to express different characteristics of overall satisfaction. For each data set the Cronbach's Alpha value is above 0.7. Transformation of the dataset is discussed in section V.

V. *Data Transformation*

Data Transformation Skewness/Kurtosis

	BEFORE TRANSFORMATION						AFTER TRANSFORMATION					
	Mean	Skewness	Skewness Z Score	Kurtosis	Kurtosis Z Score	Frequency Histogram	Mean	Skewness	Skewness Z Score	Kurtosis	Kurtosis Z Score	Frequency Histogram
Overall	7.52	-0.974	-7.730159	1.291	5.1230159		0.4687	-0.475	-3.769841	-0.51	-2.02381	
Safety	8.26	-1.448	-14.62626	1.79	9.040404		0.3238	0.344	3.4747475	-1.157	-5.843434	
Dist	7.75	-1.065	-10.75758	0.679	3.4467005		0.4102	-0.045	-0.454545	-1.145	-5.812183	
Freq	7.06	-0.705	-7.121212	-0.165	-0.837563		0.5078	-0.423	-4.272727	-0.757	-3.84264	
Transfer	7.68	-1.115	-10.72115	0.663	3.1722488		0.4183	0.019	0.1826923	-1.08	-5.167464	
Driver	7.94	-1.141	-11.41	0.911	4.5778894		0.3826	0.063	0.63	-1.193	-5.994975	
Comfort	7.62	-1.132	-11.43434	1.147	5.7929293		0.4412	-0.186	-1.878788	-0.847	-4.277778	
Fare	7.2	-0.918	-9.089109	0.214	1.0594059		0.4833	-0.29	-2.871287	-0.864	-4.277228	
Ticket	8.72	-2.059	-20.18627	4.395	21.544118		0.253	0.782	7.6666667	-0.43	-2.107843	
RideTime	7.25	-0.938	-9.474747	0.442	2.2323232		0.4908	-0.349	-3.525253	-0.614	-3.10101	
Driving	7.71	-1.034	-10.34	0.772	3.879397		0.4235	-0.129	-1.29	-1.051	-5.281407	
DirectRte	8	-1.198	-12.10101	1.208	6.1010101		0.376	0.07	0.7070707	-1.176	-5.939394	
Crowded	6.71	-0.72	-7.272727	-0.024	-0.121212		0.5567	-0.595	-6.010101	-0.197	-0.994949	
OnTime	7.13	-0.901	-9.10101	0.271	1.3686869		0.5016	-0.33	-3.333333	-0.686	-3.464646	
Info	7.74	-1.224	-11.8835	1.163	5.6731707		0.4035	-0.003	-0.029126	-1.153	-5.62439	
	Kurtosis and Skewness: Statistic/Std. Error=z score					-1.96<Z>1.96						

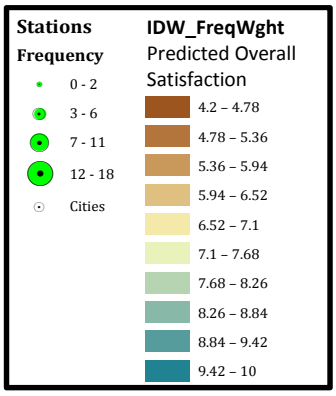
Data transformation is used in the case where the data is not normally distributed. There are many statistical tests which assume a normal distribution in the data. Two popular forms of data transformation are the use of square roots and natural log. The above data compares the results of the raw data distribution for the 14 quality parameters before and after a natural log transformation.

First data is checked for normality focusing on Kurtosis and Skewness: the kurtosis or Skewness statistic is divided by its standard error producing a z score, if -1.96<Z>1.96 then the data is normally distributed. The z score data is reported in the tables, a gray filled cell highlights normal distribution. In addition to the z score, a simple histogram is the best method for checking normality, unfortunately, none of the data produces a normal bell curve before or after transformation.

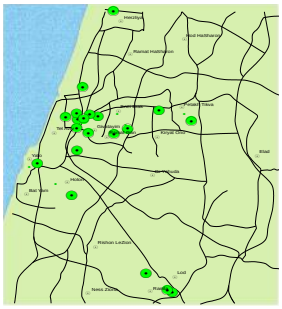
A log transformation is computed by log10(variable) which produces a transformed data set. In the case of a negative Skewness statistic an alternate formula of log10([maxvalue+1]-variable) is used. In all cases Skewness was negative and the maxvalue is always 10.

Unfortunately the log transformation had only the effect of reversing the distribution. A square root transformation was also tested with even less acceptable results such that most tests will utilize the raw data with an understanding that it is not normally distributed.

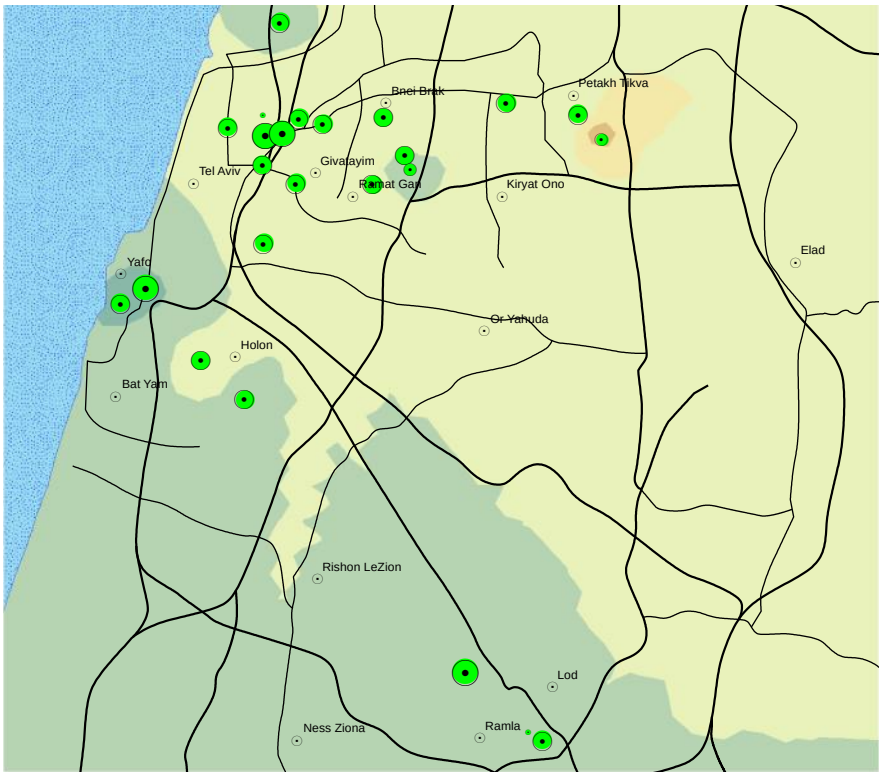
VI. *Central Station Data*



Average Overall Satisfaction



Survey Station location and Frequency; Interpolation of overall PT satisfaction



Map and table of central region stations. These stations represent 48% of the total survey data. There are 25 stations with an average frequency near 10. The map shows station frequency with symbol size.

An inverse distance weighting interpolation weighted by frequency was used to predict overall satisfaction scores for the whole region. A small minimap shows the color representative of the average overall satisfaction for the central reason which had a value of 7.42.

This map is a loose representation as so few measured data points cannot provide an accurate prediction however it shows the power of such mapping for future use.

In the case of this research, the focus will be on the stations more closely related with Tel Aviv, meaning the Ramla stations will most likely not be covered. This map also has a direct connection with the contracts expected to be included in this research making this map an important detail in the analysis of demand.

ID#	Station Name	City	Frequency
1	Saret/Karen Yasod	Holon	10
25	Etzl	Tel Aviv	10
26	Moshe Dyan	Tel Aviv	10
27	Jerusalem Blvd.	Yafo	16
28	Shivtey Israel	Yafo	9
29	Ichilov Hospital	Tel Aviv	15
30	Azrieli Mall	Tel Aviv	10
31	Kogel/Sokolov	Holon	11
32	Iben Gvirol/Arlozorov	Tel Aviv	10
33	Jabotinsky/Bialik	Ramat Gan	10
34	Ha-Yetsira	Ramat Gan	10
35	Rabbi Akiva	Bnei Brak	10
36	Khazon Ish	Bnei Brak	10
37	Petakh Tikva Market	Petakh Tikva	10
38	Ein Ganim/A.D. Gordon	Petakh Tikva	5
39	Belinson Hospital	Petakh Tikva	10
40	Yarden/Negba	Ramat Gan	10
42	Central Station	Ramla	10
43	Hertzl	Ramla	18
44	Hertzl/Frenkel	Ramla	
47	Hei be-Iyar	Tel Aviv	1
48	Esra	Bnei Brak	6
49	HaMa'apilim	Herzliya	2
91	Einstein	Tel Aviv	11
202	Arlozorov	Tel Aviv	14

VII. *Compare Means*

Selected Compare Means Significance

		Scr_Overall	Scr_Safety	Scr_Dist	Scr_Freq	Scr_Transfer	Scr_DriverBhv	Scr_Comfort	Scr_Fare	Scr_Ticket	Scr_RideTime	Scr_Driving	Scr_DirectRte	Scr_Crowded	Scr_OnTime	Scr_Info
Region	F	2.117	2.376	0.288	1.515	2.311	4.808	3.237	5.293	5.155	0.22	0.557	4.047	6.062	3.67	2.006
	Sig.	0.098	0.069	0.834	0.21	0.075	0.003	0.022	0.001	0.002	0.883	0.643	0.007	0	0.012	0.112
Trip_Obj	F	1.559	0.929	1.299	2.179	2.231	0.719	2.395	2.678	0.595	1.145	0.681	3.059	3.437	1.902	1.385
	Sig.	0.158	0.474	0.255	0.043	0.039	0.635	0.027	0.014	0.734	0.335	0.665	0.006	0.002	0.078	0.218
Gender	F	0.043	1.951	0.158	1.486	0.004	0.545	4.669	2.093	5.933	0.21	0.584	0.547	1.632	0	1.609
	Sig.	0.836	0.163	0.692	0.223	0.952	0.461	0.031	0.149	0.015	0.647	0.445	0.46	0.202	0.99	0.205
Trip_Type	F	0.032	0.665	6.118	5.541	0.672	1.305	0.043	2.248	1.143	0.914	3.908	0.081	2.623	4.09	0.526
	Sig.	0.857	0.415	0.014	0.019	0.413	0.254	0.837	0.134	0.286	0.34	0.049	0.776	0.106	0.044	0.469
Freq_Chng	F	1.462	3.213	1.561	4.455	2.349	1.733	2.384	5.495	1.973	1.299	1.128	3.545	0.299	1.181	1.323
	Sig.	0.225	0.023	0.198	0.004	0.072	0.159	0.068	0.001	0.117	0.274	0.337	0.014	0.826	0.316	0.266
Trip_Freq_Grp	F	1.317	1.543	1.211	0.563	11.742	0.151	0.359	0.343	2.41	2.761	4.367	2.598	0.723	1.084	7.593
	Sig.	0.252	0.215	0.271	0.453	0.001	0.698	0.549	0.559	0.121	0.097	0.037	0.108	0.395	0.298	0.006
Age_Grp	F	0.748	1.313	1.651	0.226	0.693	0.468	1.275	5.71	0.741	1.106	0.464	0.832	1.231	2.354	2.904
	Sig.	0.56	0.264	0.16	0.924	0.597	0.759	0.278	0	0.564	0.353	0.762	0.505	0.296	0.053	0.021

- Mean values vary by independent categories at a significance of 0.05

Selected compare means showing how differences between categories in categorical variables are statistically significant when considering the 14 quality parameters. Blue highlighted cells represent parameters which are different from one category to the next at a statistically significant level.

Region- there are a number of parameters which vary by region, mainly crowdedness, fare, tickets, and directness of the route.

Trip Objective- there is variance by trip objective but on a low scale for the parameters: crowdedness, direct route and fare.

Gender - there is strong variance for only two parameters: comfort and ticket.

Trip Type - for trip type, the parameters of distance from station, frequency, reliability (on time) and driving are the most important.

Frequency Change - there is significant variance between those reporting ride frequency change for the parameters of frequency and fare.

Trip Frequency Group - significant variance exists between groups regarding transfer tickets and information availability.

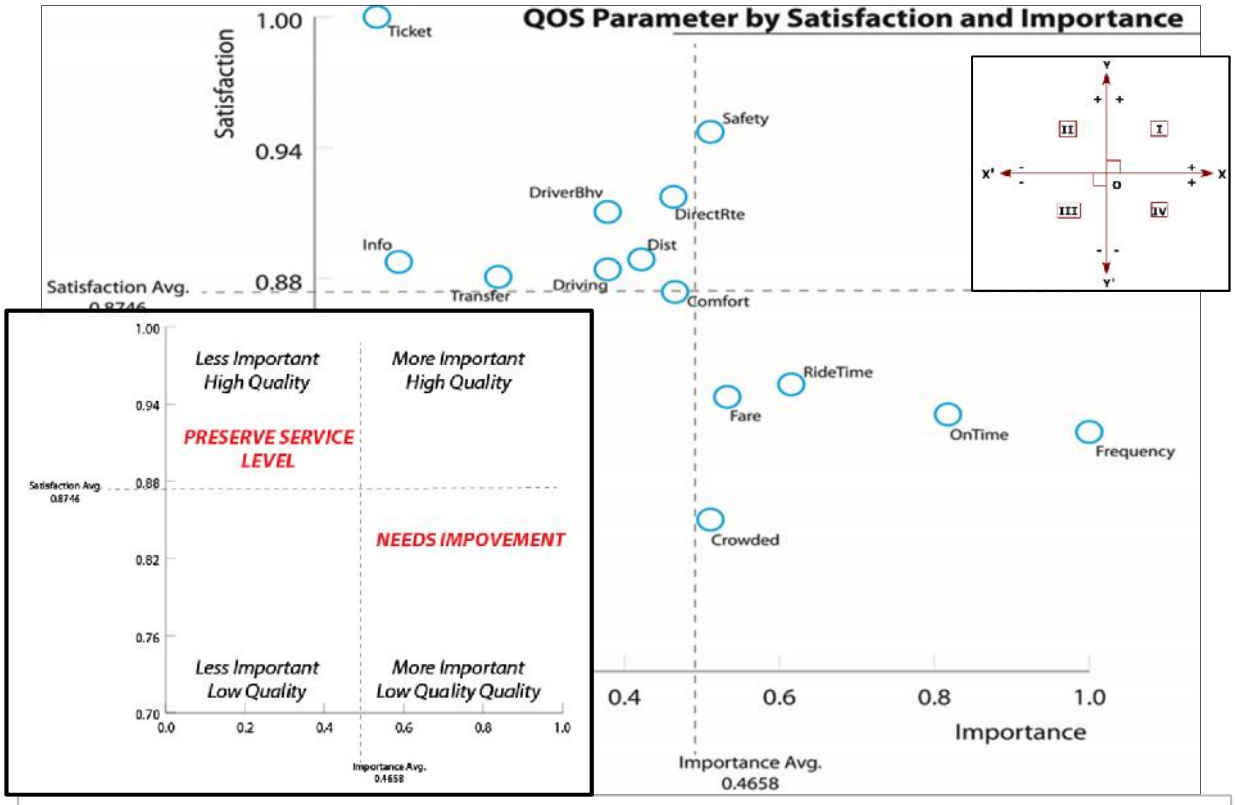
Age Group - fare has a slight variance between age groups.

Sub Conclusion -

The comparison of means shows the significance and magnitude of change of average values for different variable categories. For some variables the variance is very intuitive such as the variance in transfer tickets and information availability for trip frequency groups. Another example is the importance of objective, time sensitive parameters (frequency, reliability and distance from station) for those using the bus for inter or intracity journeys. The lack of variance in overall satisfaction shows that for the entire survey population the average is congruent.

VIII. *Importance and Satisfaction*

	Importance					Satisfaction	
	Percent of Frequency						
	I	II	III	Avg Percent	Prop Avg. Percent	Avg. Satisfaction	Prop Satisfactio
Safety	15.9	4.2	3.4	7.8	0.511	8.26	0.947
Dist	7.9	6.7	4.8	6.5	0.422	7.75	0.889
Frequency	25	13.2	7.8	15.3	1.000	7.06	0.810
Transfer	3.2	3.9	3.8	3.6	0.237	7.68	0.881
DriverBhv	5.5	7.8	4.1	5.8	0.378	7.94	0.911
Comfort	4.2	10.3	6.9	7.1	0.465	7.62	0.874
Fare	4.7	11.5	8.3	8.2	0.533	7.2	0.826
Ticket	0.8	1.9	1	1.2	0.080	8.72	1.000
RideTime	6.9	11.1	10.3	9.4	0.615	7.25	0.831
Driving	2.7	6.1	8.6	5.8	0.378	7.71	0.884
DirectRte	4.9	7.6	8.8	7.1	0.463	8	0.917
Crowded	5	6.9	11.6	7.8	0.511	6.71	0.769
OnTime	11.9	8.1	17.6	12.5	0.817	7.13	0.818
Info	1.3	0.8	2.9	1.7	0.109	7.74	0.888

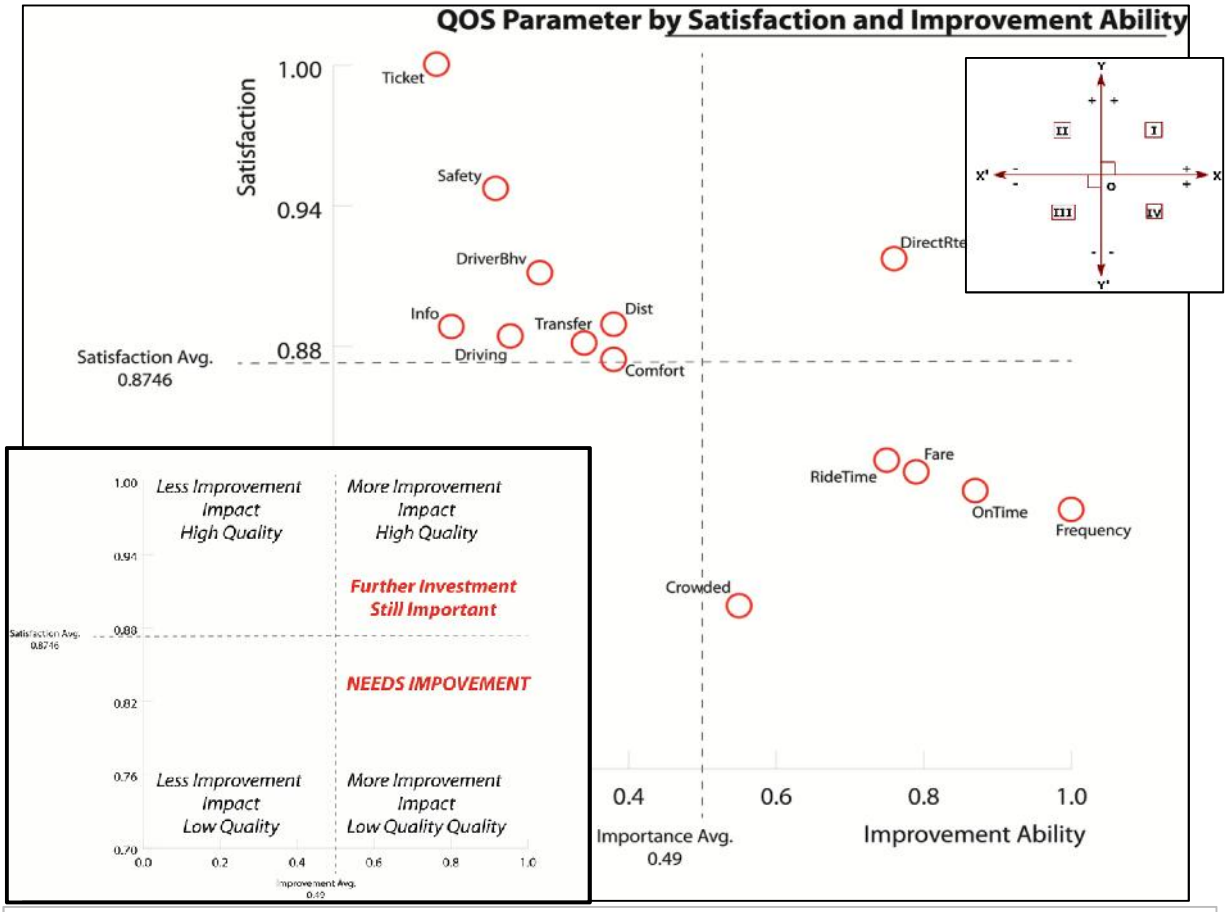


When graphing satisfaction and average importance together. In the lower righthand quadrant are most of the 'objective' parameters while more subjective ones are grouped in the upper lefthand quadrant. Quadrants 1 and 3 are nearly empty. The mini graph shows quadrant labels and that the parameters in quadrant 4 are the most important and require improvement.

IX. *Improvement and Satisfaction*

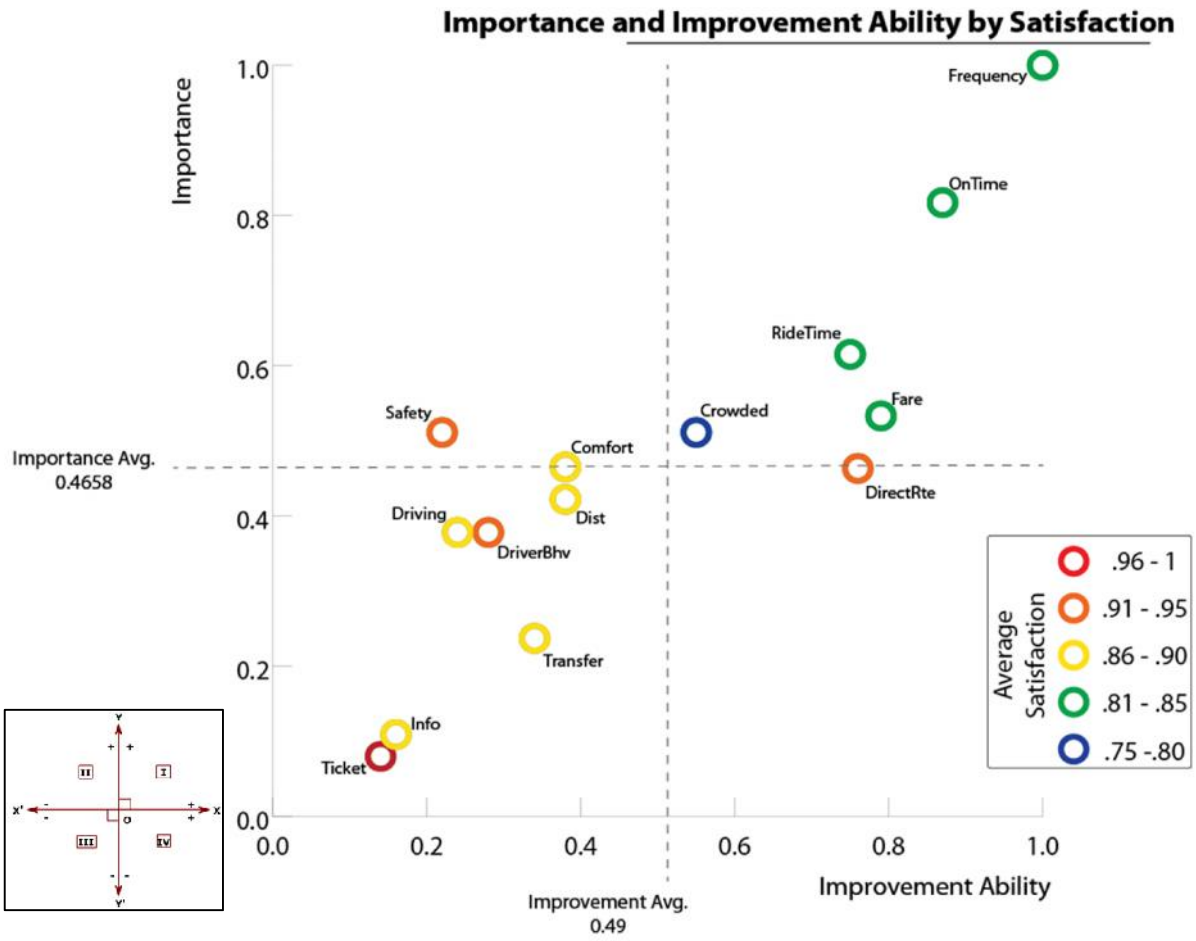
	Improvement					Satisfaction	
	Percent of Frequency						
	I	II	III	Avg. Percent	Prop Avg. Percent	Avg. Satisfaction	Prop Avg. Satisfaction
Safety	4.20	2.30	3.09	3.20	0.22	8.26	0.947
Dist	5.04	3.69	7.73	5.49	0.38	7.75	0.889
Frequency	22.27	10.60	10.82	14.56	1.00	7.06	0.810
Transfer	6.30	5.53	3.09	4.98	0.34	7.68	0.881
DriverBhv	2.52	5.07	4.64	4.08	0.28	7.94	0.911
Comfort	6.30	5.53	4.64	5.49	0.38	7.62	0.874
Fare	13.87	11.98	8.76	11.54	0.79	7.2	0.826
Ticket	1.68	2.30	2.06	2.02	0.14	8.72	1.000
RideTime	5.88	16.59	10.31	10.93	0.75	7.25	0.831
Driving	0.84	5.07	4.64	3.52	0.24	7.71	0.884
DirectRte	7.98	12.44	12.89	11.10	0.76	8	0.917
Crowded	6.30	7.83	9.79	7.98	0.55	6.71	0.769
OnTime	15.55	8.76	13.92	12.74	0.87	7.13	0.818
Info	1.26	2.30	3.61	2.39	0.16	7.74	0.888
Valid Percentages without Frequency of 0 calculated							

Prop - proportional values, each value is divided by the highest value in order to fit a scale of 1-10



In a similar trend to the previous graph, the impact of improvements of quality parameter are perceived as higher for parameters with lower average satisfaction. Again the objective parameters are in the fourth quadrant, labeled in the mini graph as needing improvement.

X. *Importance, Improvement and Satisfaction*



A novel graph showing the relationship between PT rider defined overall satisfaction, importance and improvement ability for the 14 QOS parameters. The parameters in the third quadrant are less important, and have less improvement ability but are also ranked highest in satisfaction. Opposite this in the first quadrant are parameters which are important and have high improvement ability but relatively low satisfaction scores. It appears that as overall satisfaction for a parameter improves, its perceived importance and improvement ability decreases. It is also important to note that again the objective parameters of frequency, reliability, ride time and fare are grouped while more subjective parameters are grouped in the third quadrant.

XI. *Linear Regression Models for Predicting Overall Satisfaction*

Pearson Correlations for dependent variable: Overall Satisfaction

Scr_Overall															
Pearson Correlation															
	Scr_Overall	Safety	Dist	Frequency	Transfer	DriverBhv	Comfort	Fare	Ticket	RideTime	Driving	DirectRte	Crowded	OnTime	Info
Correlations	1	0.282	0.334	0.541	0.509	0.343	0.433	0.328	0.291	0.48	0.479	0.467	0.407	0.515	0.436
Sig. (2-tailed)		0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	372	368	369	370	316	367	368	352	350	365	363	367	365	367	353

Model Summary of Linear regression

Linear Regression Model Summaries, dependant variable: Overall Satisfaction														
	Safety	Dist	Frequency	Transfer	DriverBhv	Comfort	Fare	Ticket	RideTime	Driving	DirectRte	Crowded	OnTime	Info
Rsquare	0.08	0.112	0.292	0.259	0.118	0.187	0.108	0.085	0.231	0.23	0.218	0.166	0.265	0.19
Residual Sum of Sqr	1234.278	1197.28	953.997	902.752	1149.704	1091.647	1134.485	1189.129	996.924	993.144	1048.512	1075.401	970.861	981.298
Standardized Coefficient	0.282	0.334	0.541	0.509	0.343	0.433	0.328	0.291	0.48	0.479	0.467	0.407	0.515	0.436
Coefficient Sig.	0	0	0	0	0	0	0	0	0	0	0	0	0	0

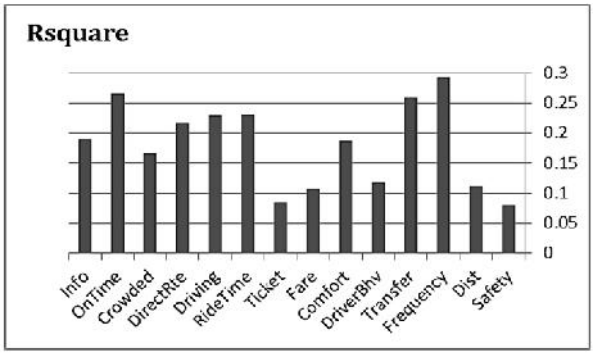
Correlations - correlations show the strength of the connection between two variables, in this case the table shows that all of the QOS parameters have a positive correlation with high levels of significance. The highest correlations are for frequency and reliability with variables such as transfer and ride time close behind. A correlation of 0.500 is considered moderately strong.

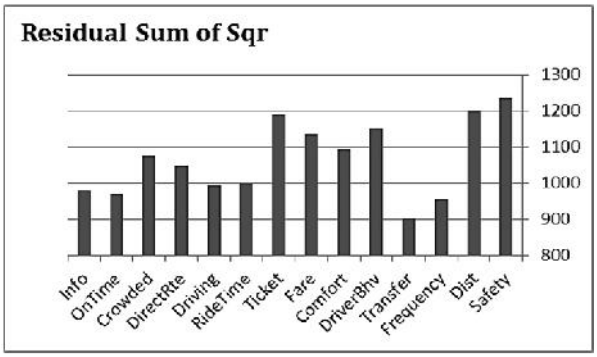
Regression Summaries - linear regression expresses the ability of an independent variable to predict the dependent variable. In this case the dependent variable is overall satisfaction.

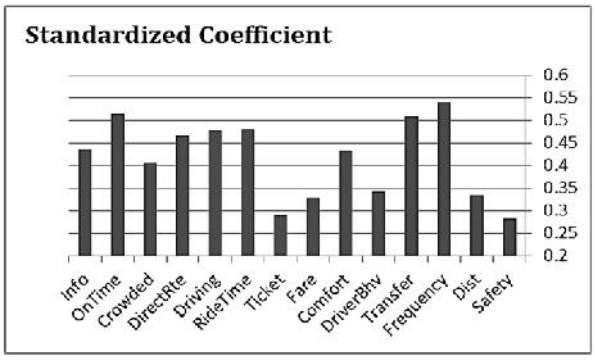
Rsquare - expresses the ability of the prediction to explain the variance in measured data points. Frequency and reliability have the greatest ability to predict satisfaction in a limited capacity.

Residual Sum of Sqr - the residual value expresses the difference between the predicted data and the measured, the lower the residual value, the more accurate the prediction.

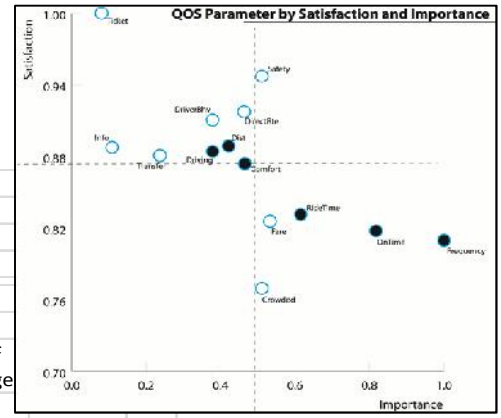
Sub Conclusion - objective factors have the strongest ability to predict overall satisfaction in a single variable regression model. The residual sum of squares is very high for all values but consistent for each model. Overall the Rsquare values are low meaning single parameters are







XII. *Multi Variable Regression*



Plot of parameters as defined by stepwise multi variable regression

Stepwise Multi Variable Regression

Dependent Variable: Scr_Overall

Independent Variables: All Parameter Scores

	Model Summary			Change Statistics		
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	.568a	0.323	1.585	0.323	131.994	0
2	.648b	0.42	1.469	0.097	46.215	0
3	.683c	0.466	1.412	0.046	23.838	0
4	.700d	0.491	1.382	0.024	13.141	0
5	.712e	0.507	1.361	0.017	9.346	0.002
6	.719f	0.516	1.351	0.009	5.019	0.026

a Predictors: (Constant), Scr_Freq

b Predictors: (Constant), Scr_Freq, Scr_RideTime

c Predictors: (Constant), Scr_Freq, Scr_RideTime, Scr_Comfort

d Predictors: (Constant), Scr_Freq, Scr_RideTime, Scr_Comfort, Scr_Driving

e Predictors: (Constant), Scr_Freq, Scr_RideTime, Scr_Comfort, Scr_Driving, Scr_OnTime

f Predictors: (Constant), Scr_Freq, Scr_RideTime, Scr_Comfort, Scr_Driving, Scr_OnTime, Scr_Dist

g Dependent Variable: Scr_Overall

Coefficients

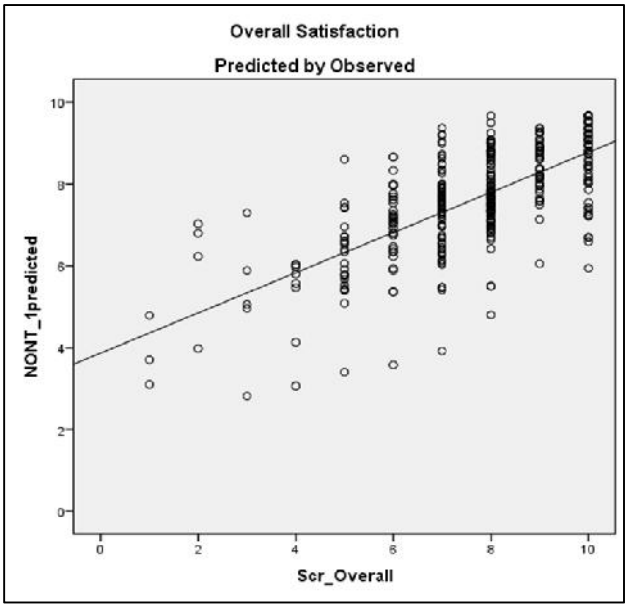
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
6	(Constant)	1.137	0.416		2.73	0.007
	Scr_Freq	0.166	0.047	0.205	3.533	0
	Scr_RideTime	0.167	0.041	0.207	4.081	0
	Scr_Comfort	0.143	0.049	0.15	2.944	0.004
	Scr_Driving	0.123	0.044	0.139	2.767	0.006
	Scr_OnTime	0.147	0.045	0.182	3.228	0.001
	Scr_Dist	0.107	0.048	0.115	2.24	0.026

a Dependent Variable: Scr_Overall

Multi Variable Regression - this regression model used the stepwise method entering variables which had a significant positive effect on the Rsquare value. Six parameters were entered achieving a Rsquare value of .516 which is moderately strong. Individually the parameter coefficients are weak but together the model provides an acceptable prediction of half of the variance in the observed data of overall satisfaction.

Included Parameters:

Frequency
Ride Time
Comfort
Driving
On Time

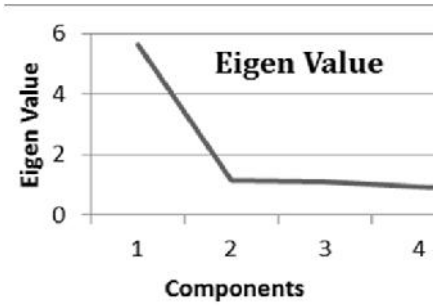


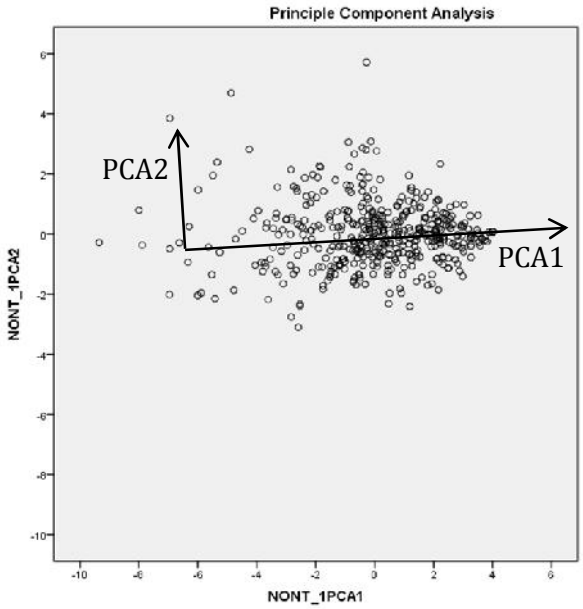
XIII. *Principle Component Analysis*

KMO and Bartlett's Test			
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.			0.878
Bartlett's Test of Sphericity	Approx. Chi-Square		2311.081
		df	91
		Sig.	0

Total Variance Explained			
Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.618	40.128	40.128
2	1.134	8.101	48.229
3	1.09	7.787	56.016

Component Matrix			
	Component		
	1	2	3
Scr_Safety	0.534	0.275	0.33
Scr_Dist	0.594	-0.001	0.133
Scr_Freq	0.661	-0.441	-0.1
Scr_Transfer	0.662	-0.353	0.194
Scr_DriverBhv	0.637	0.303	0.188
Scr_Comfort	0.666	0.227	-0.427
Scr_Fare	0.538	0.413	-0.301
Scr_Ticket	0.55	0.157	0.222
Scr_RideTime	0.676	0.03	0.072
Scr_Driving	0.677	0.262	0.306
Scr_DirectRte	0.693	-0.093	0.259
Scr_Crowded	0.631	0.154	-0.578
Scr_OnTime	0.704	-0.348	-0.238
Scr_Info	0.613	-0.422	-0.025





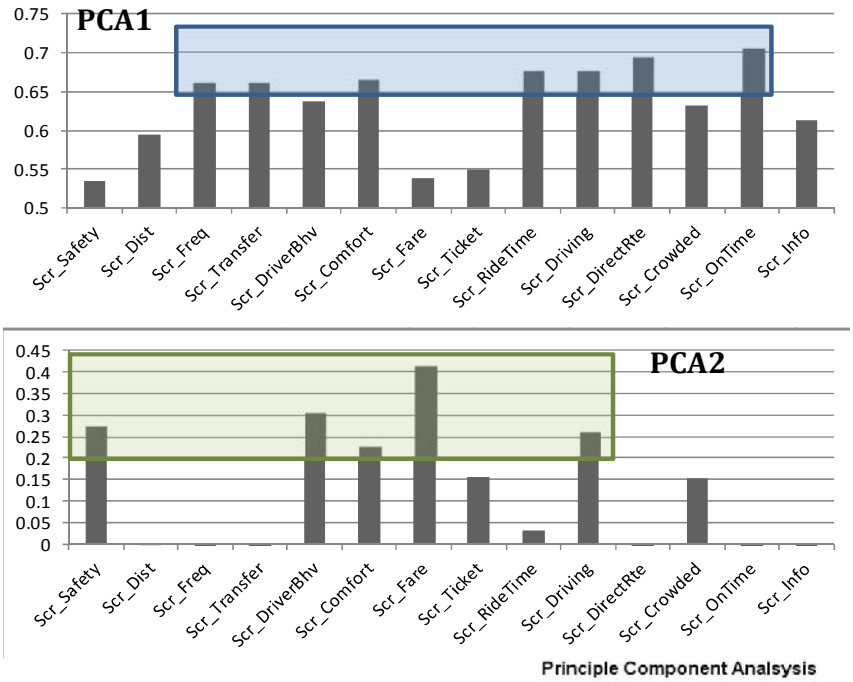
Principle Component Analysis - this is a popular which combines multiple variables, in this case QOS parameters, into single components based on their value in the component matrix. The amount of components extracted by the model is based on Eigen value which expresses how well the amount of components explain the variance of the individual cases. While many components can be extracted, a two component model is intuitive and therefore favorable for data interpretation.

KMO test - the high value means that the variables are correct for performing PCA calculations.

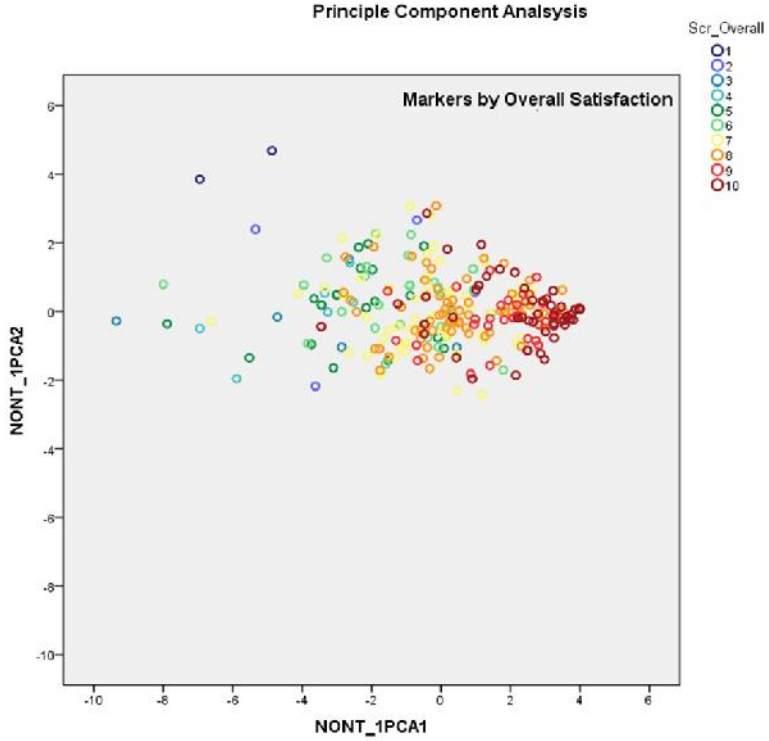
Bartlett's test - a significance level of .000 means that there is some level of correlation between the variables.

Variance Explained - two components explain approximately 48% of the data, additional components have little effect on variance explained.

XIV. *Principle Component Analysis Continued*



Component Matrix		
	Component	
	1	2
Scr_Safety	0.534	0.275
Scr_Dist	0.594	-0.001
Scr_Freq	0.661	-0.441
Scr_Transfer	0.662	-0.353
Scr_DriverBhv	0.637	0.303
Scr_Comfort	0.666	0.227
Scr_Fare	0.538	0.413
Scr_Ticket	0.55	0.157
Scr_RideTime	0.676	0.03
Scr_Driving	0.677	0.262
Scr_DirectRte	0.693	-0.093
Scr_Crowded	0.631	0.154
Scr_OnTime	0.704	-0.348
Scr_Info	0.613	-0.422



Component Matrix - on the previous page and this are component matrices which show how the parameters are sorted into the two components. The bar graphs further show the prominence of certain parameters in each component.

When the PCA1/2 variables are plotted and marked by overall satisfaction, there is a clear trend along the PCA1 line. As PCA1 values increase, overall satisfaction grows, this shows that the variables included in this component have the ability to group respondents by their level of overall satisfaction. The second component contributes less to the discussion.

XV. *Multi Variable Regression based on PCA1/2*

Enter Method Multi Variable Linear Regression Based on PCA Components

PCA1

Dependent Variable: Scr_Overall

Independent Variables:

Scr_OnTime, Scr_Driving, Scr_Comfort, Scr_DirectRte, Scr_RideTime, Scr_Transfer, Scr_Freq

	Model Summary				Change Statistics			
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change	Residual Sum of Sqr
1	.717a	0.513	0.502	1.361	0.513	45.218	0	586.494

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	1.158	0.394		2.94	0.004
Scr_Freq	0.17	0.048	0.21	3.572	0
Scr_Transfer	0.07	0.044	0.088	1.59	0.113
Scr_Comfort	0.118	0.045	0.123	2.648	0.009
Scr_RideTime	0.156	0.042	0.189	3.748	0
Scr_Driving	0.158	0.041	0.179	3.814	0
Scr_DirectRte	0.062	0.047	0.071	1.32	0.188
Scr_OnTime	0.118	0.044	0.143	2.704	0.007

PCA2

Dependent Variable: Scr_Overall

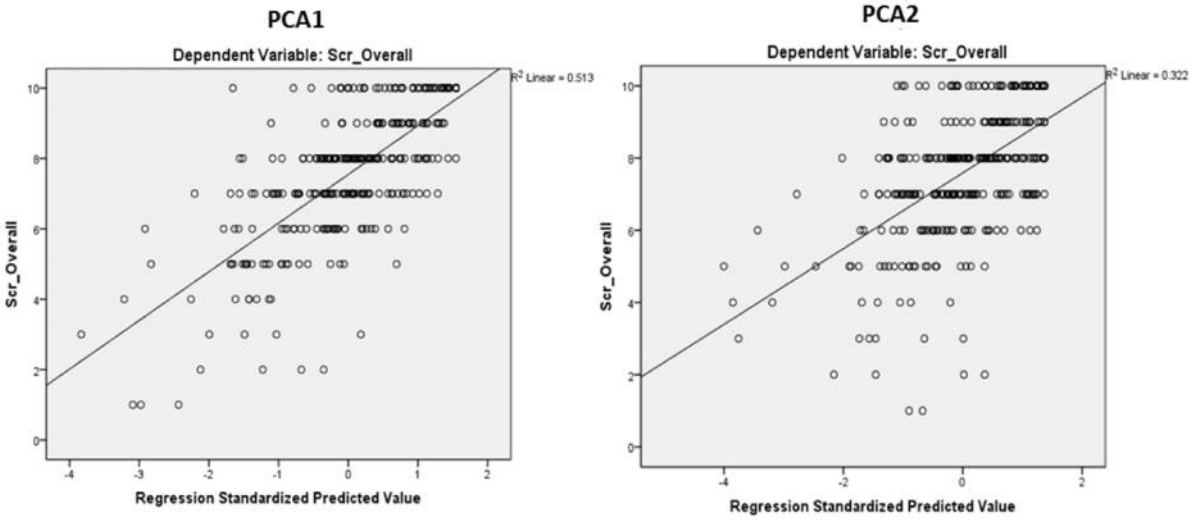
Independent Variables: Scr_Driving, Scr_Fare, Scr_Safety, Scr_Comfort, Scr_DriverBhv

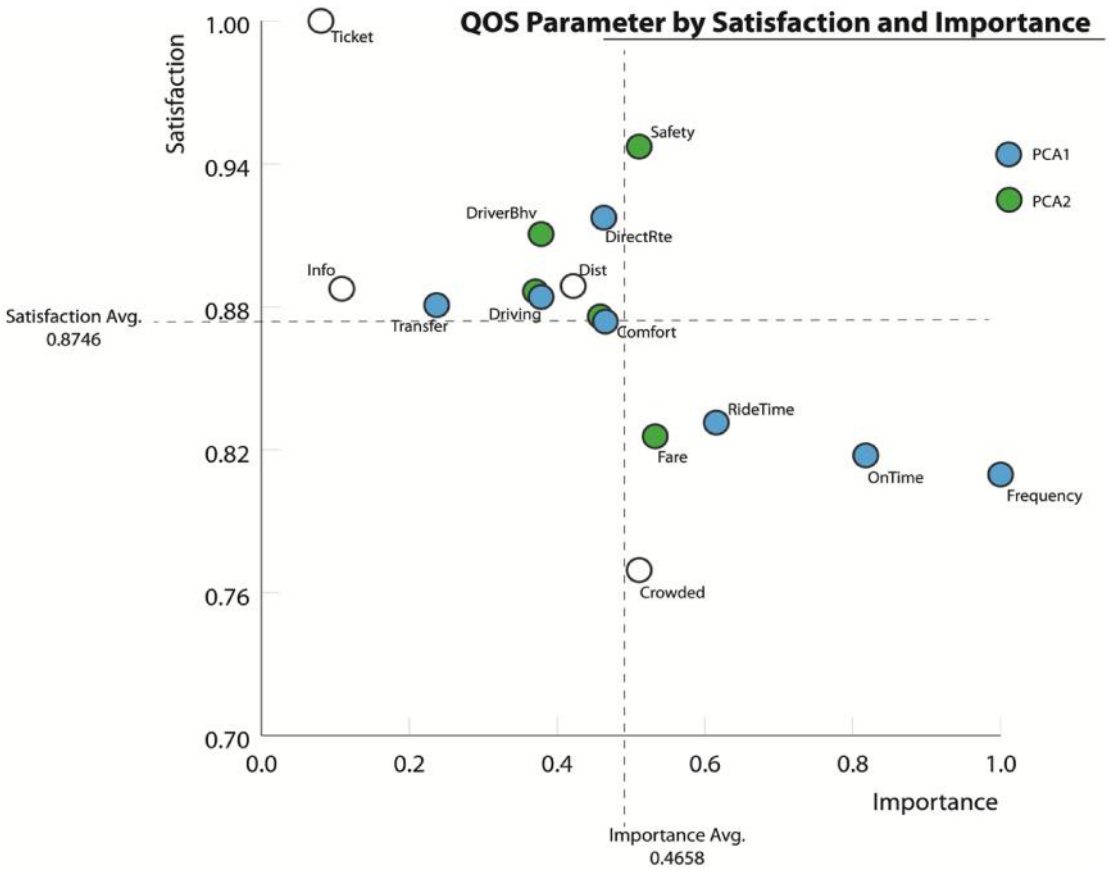
	Model Summary				Change Statistics			
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change	Residual Sum of Sqr
1	.568a	0.322	0.312	1.533	0.322	31.786	0	784.754

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	2.332	0.466		5.006	0
Scr_Safety	0.032	0.048	0.034	0.671	0.503
Scr_DriverBhv	0.056	0.055	0.061	1.025	0.306
Scr_Comfort	0.264	0.05	0.281	5.236	0
Scr_Fare	0.065	0.04	0.084	1.636	0.103
Scr_Driving	0.252	0.051	0.29	4.928	0

XVI. *Multi Variable Regression based on PCA1/2 Continued*





A enter method multi variable regression model shows a dual model construct which explains half of the variance when predicting overall satisfaction. This regression analysis is similar to the first but differs in two ways. First, the enter method included all of the entered independent variables as defined by the PCA test, together these parameters explain 51% and 32% of the data. Second, this model shows how the PCA selected parameters in groups, one more objective and the other subjective. This has been a common trend in the data up to this point and further supports which parameters are most important to PT users.