

---

*COMP 3839 Data Quality Improvement*

*Vijay Laxmi Tyagi (A01317295)*

*Iteration 3*

---

## Contents

Overview of Project .....	3
Initial Assessment .....	4
LicenceRSN : Duplicates .....	4
Business Name : Null.....	5
PostalCode: Invalid data Format.....	6
Status : Invalid Data .....	7
Province : 2-Capitalized Letter Standard Not Followed.....	7
FeePaid: Null Values.....	8
.....	9
SME Review of Initial Assessment .....	10
LicenceRSN : Duplicates (High) .....	10
Business Name : Null (Moderate) .....	10
PostalCode: Invalid data Format (Moderate) .....	10
Status : Invalid Data (low) .....	10
Province : 2-Capitalized Letter Standard Not Followed (Moderate) .....	10
FeePaid: Null Values (moderate) .....	10
Further Research for the SME.....	11
SME Review of Further Research.....	11
SME Suggests Some DQ Rules.....	11
LicenceRSN.....	11
Business Name .....	11
PostalCode .....	11
Status .....	11
Province .....	12
FeePaid.....	12

## Overview of Project

Dataset used for quality Analysis is of business licence record for the year 2015. City of Vancouver is the owner of this Data Set. Total Number of records in this data set are 60938. Licence RSN is a Unique Identifier and all the Licence RSN are Distinct. It's an open source dataset and the link for the same is

<https://opendata.vancouver.ca/explore/dataset/business-licences>

### Columns and their brief Description is as Follows:-

Column Name	Description
FolderYear	Represents the year in which Business Licence Number is issued. In our Dataset it's 15 which signifies year 2015
LicenceRSN	Identifier for each business licence. It's a Unique field generated by the system
LicenceNumber	This is a 9 character field consisting of a two digit number indicating the year issued followed by a hyphen and a six digit system-generated number
LicenceRevisionNumber	A 2 digit field generated by the system that represents the licence version
BusinessName	Name of the owner of Business
BusinessTradeName	Trade Mark Name of the Business
Status	Current status of the Business whether it's Issued, Pending, Cancelled or gone out of business etc.
IssueDate	Date when the business licence was issued
ExpireDate	The date that the business licence expires
BusinessType	Type/ Category which broadly describes the Business
BusinessSubtype	Sub-category of the main business type
Unit	This is an alphanumeric field. Represents official space
UnitType	Description of a location i.e., Suite, Apartment (Apt.), etc
House	The number assigned to an address where the business is located.
Street	The name of the street where the business is located.
City	Name of the city where the business is located
Province	Name of the province where the business is located
Country	First two characters of name of the country where the business is located
PostalCode	Postal code attached to the address of a business
LocalArea	Area where Business is Located
NumberOfEmployees	Total Number of Employees in Business
FeePaid	Total amount of licence fee paid in Canadian dollars
ExtractDate	Date when data was extracted from source data system.

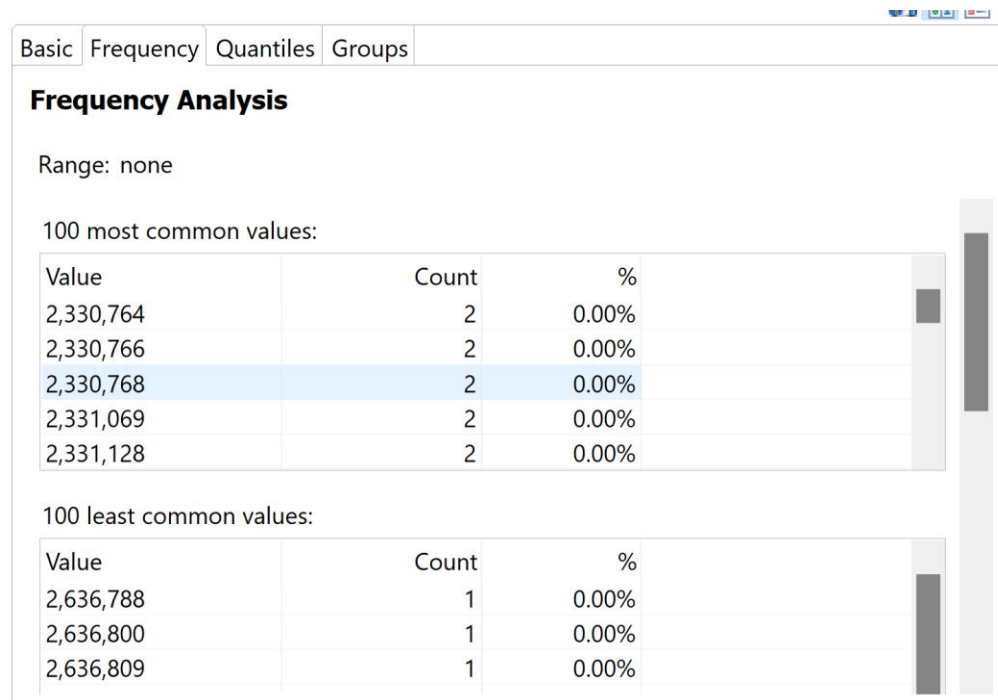
Column Name	Description
Geom	Spatial representation

In the first section i.e. initial assessment, me being the technical analyst described assessment of Business licences data of city of Vancouver. For assessment I have used data quality Analyzer tool of Ataccama. In the next section of this Project report, SME i.e. Business Analyst in City of Vancouver has reviewed the Assessments made. She had analysed the data quality issues and give her feedback and suggestions if any. Then I did further research based on SME's feedback. Then in the last section SME reviewed the further Research and suggest DQ rules.

## Initial Assessment

### LicenceRSN : Duplicates

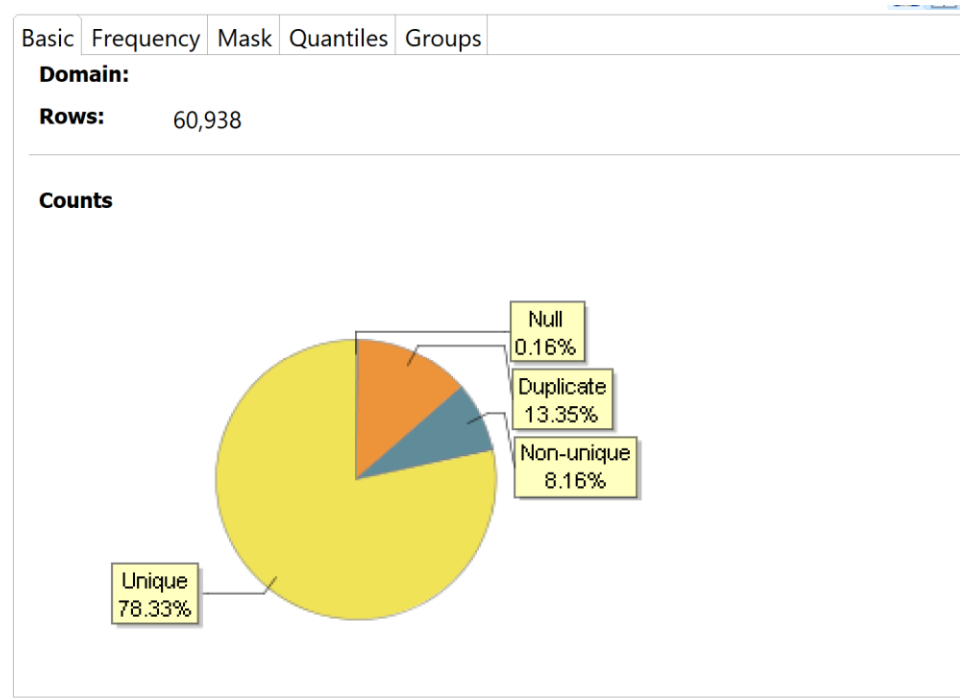
LicenceRSN is a unique Identifier field and all the LicnceRSN Number should be Unique/Distinct but we have observed there are multiple Records with same value. As LicenceRSN is a unique identifier it will be used in searching, inserting, updating and deleting data conditions. I am attaching screenshots which will show that this attribute has duplicate values.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
853	15	2330758	15-10282400		Sandra Wong Kennet	Issued		2014-11-3	2015-12-31	Apartment House				8738	French St	Vancouver BC	CA	VI	
854	15	2330759	15-10282700		Mighty Atom Holding	Issued		2014-12-0	2015-12-31	Apartment House				203	E 6TH AV	Vancouver BC	CA	VI	
855	15	2330760	15-10283000		EMK Propi Park Bay A	Issued		2015-01-1	2015-12-31	Apartment House				1010	GILFORD S	Vancouver BC	CA	VI	
856	15	2330761	15-10283100		EMK Propi Kathleen C	Issued		2015-01-1	2015-12-31	Apartment House				1555	HARWOOD	Vancouver BC	CA	VI	
857	15	2330762	15-10283500		Wendy Lo Park Avon	Issued		2014-12-2	2015-12-31	Apartment House				8741	CARTIER S	Vancouver BC	CA	VI	
858	15	2330763	15-10284100		Radio City Radio City	Issued		2014-12-2	2015-12-31	Apartment House				1031	Burnaby S	Vancouver BC	CA	VI	
859	15	2330764	15-10285100		Derlan Holdings Ltd	Issued		2014-11-2	2015-12-31	Apartment House				1835	Comox St	Vancouver BC	CA		
860	15	2330764	15-10285700		E Alsilmi & Group Ltd	Gone Out of Business				Apartment House				3708	Knight St	Vancouver BC	CA	VI	
861	15	2330765	15-10285900		(Rudy Labuhn)	Issued		2014-11-2	2015-12-31	Apartment House				1353	Barclay St	Vancouver BC	CA	VI	
862	15	2330766	15-10286000		Project 24 Garden Ap	Issued		2014-12-0	2015-12-31	Apartment House				1503	GRANDVIE	Vancouver BC	CA		
863	15	2330766	15-10286200		Camelia Court Ltd	Issued		2014-12-0	2015-12-31	Apartment House				525	W 13th Av	Vancouver BC	CA	VI	
864	15	2330767	15-10286300		Skyline Apartments L	Issued		2014-12-0	2015-12-31	Apartment House				1139	Barclay St	Vancouver BC	CA	VI	
865	15	2330768	15-10286400		Laurentian Towers H	Issued		2014-12-0	2015-12-31	Apartment House				1636	Haro St	Vancouver BC	CA	VI	
866	15	2330768	15-10286700		Alex Yanchuk & Doro	Issued		2014-11-0	2015-12-31	Apartment House				228	W 19TH A	Vancouver BC	CA	VI	
867	15	2330769	15-10286800		Hampton Royale Apt	Issued		2015-02-0	2015-12-31	Apartment House				1650	Barclay St	Vancouver BC	CA	VI	
868	15	2330770	15-10287200		Beach Tow Douglas H	Issued		2014-12-1	2015-12-31	Apartment House				1600	BEACH AV	Vancouver BC	CA	VI	
869	15	2330771	15-10287400		English Bay Towers Ir	Issued		2014-12-1	2015-12-31	Apartment House				1651	HARWOOD	Vancouver BC	CA	VI	
870	15	2330772	15-10287500		(Natasha Taylor)	Issued		2014-12-0	2015-12-31	Apartment House				3223	FINDLAY S	Vancouver BC	CA	VI	

### Business Name : Null

This field has Names of the Businesses and Each and Every Business which is functional should have one valid Name. We have observed that 99 Businesses have Null Value. Which is very high. Best part is these none of these 99 businesses have licence status issued.



Basic	Frequency	Mask	Quantiles	Groups
<b>Frequency Analysis</b>				
Range: none				
100 most common values:				
Value	Count	%		
NULL	99	0.16%		
Imperial Parking Ca...	103	0.17%		
Starbucks Coffee Ca...	97	0.16%		
City of Vancouver	56	0.09%		
Advanced Parking S...	54	0.09%		
100 least common values:				
Value	Count	%		
Zhi Yang Chen (Zhi ...	1	0.00%		
Zhi Ying Cai (Zhi Cai)	1	0.00%		
Zhi Ying Li (Zhi Li)	1	0.00%		

### PostalCode: Invalid data Format

Vancouver Business Licences Record has Canada's and USA's Postal codes. As per Canada valid format for Postal code is 3 character followed by a space and then again 3 characters and might be either five digits or five digits followed a hyphen (dash) and another four digits (12345-1234). Whereas we have observed Multiple Records with invalid format.

Basic	Frequency	Mask	Quantiles	Groups
<b>Mask</b>				
Mask: characters: [:letter:] -> L,[:digit:] -> D				
Value	Count	%		
NULL	27,035	44.36%		
LDL DLD	33,319	54.68%		
LDLDLD	477	0.78%		
LDL DLD	17	0.03%		
LDL DLL	14	0.02%		
DDDDD	12	0.02%		
LLL LLLLLL	10	0.02%		
LDL DDD	6	0.01%		
LDL DDL	5	0.01%		
L/L	4	0.01%		
LDD DLD	4	0.01%		
LL	4	0.01%		

**Status : Invalid Data**

Status Column has 3 invalid entries

“ Ai Jia” and “Physiotherapy and Rehabilitation Center”

Status of the Business should be either Issued, Pending, Cancelled, gone out of business etc or something which tells us about proper functioning status of the Business.



Basic	Frequency	Domains	Mask	Quantiles	Groups
-------	-----------	---------	------	-----------	--------

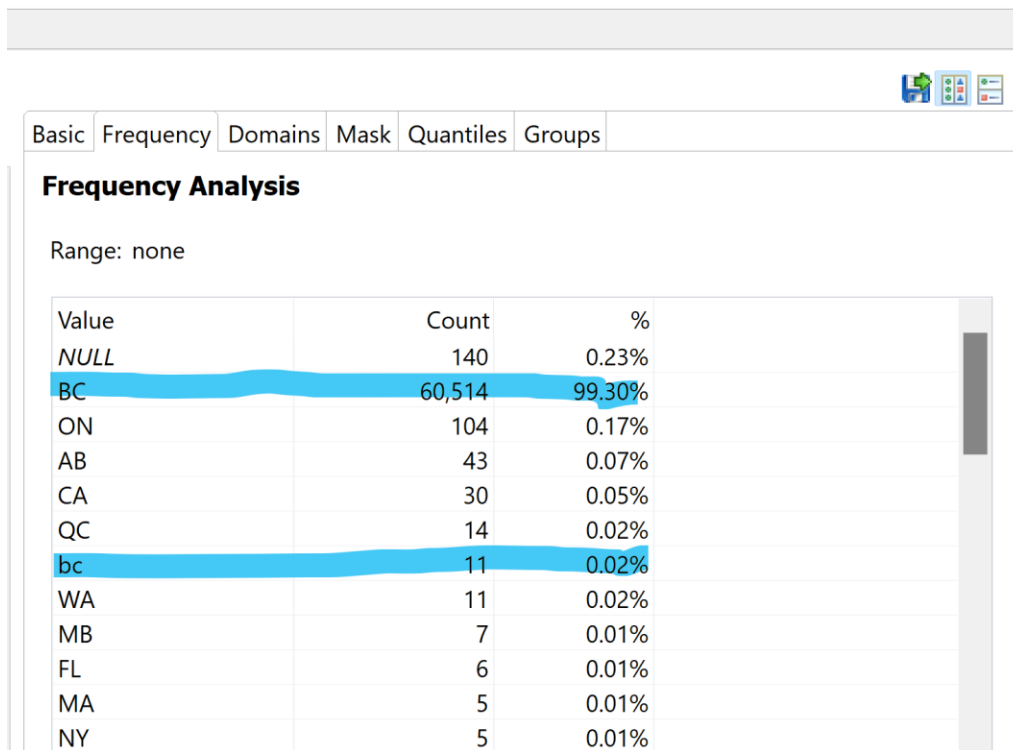
**Frequency Analysis**

Range: none

Value	Count	%
NULL	1	0.00%
Issued	48,130	78.98%
Gone Out of Business	5,480	8.99%
Pending	5,026	8.25%
Inactive	1,441	2.36%
Cancelled	857	1.41%
Ai Jia"	2	0.00%
Physiotherapy & Re...	1	0.00%

**Province : 2-Capitalized Letter Standard Not Followed**

Observed duplicate Entries as standard Naming Convention is not Used. For e.g BC( British Columbia) has been Entered as BC and bc which is actually same value but considered as different because of used Naming Convention.



Basic Frequency Domains Mask Quantiles Groups

**Frequency Analysis**

Range: none

Value	Count	%
NULL	140	0.23%
BC	60,514	99.30%
ON	104	0.17%
AB	43	0.07%
CA	30	0.05%
QC	14	0.02%
bc	11	0.02%
WA	11	0.02%
MB	7	0.01%
FL	6	0.01%
MA	5	0.01%
NY	5	0.01%

### FeePaid: Null Values

This field should contain some numeric Value as fees paid . We observed some null values. Fees value should be some numeric value if the licence status is issued but we have observed fees paid to be null even if status is being issued.

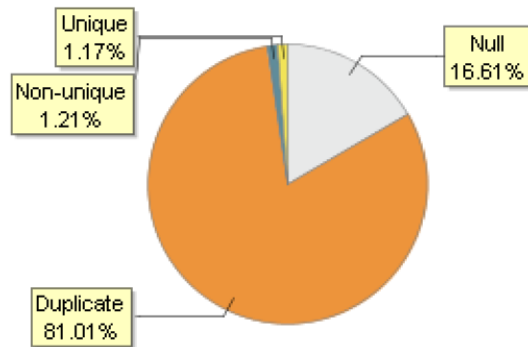


Basic Frequency Quantiles Groups

Domain:

Rows: 60,938

Counts



Type	Count	%
Null	10,120	16.61%
Non-null	50,818	83.39%
Duplicate	49,364	81.01%
Distinct	1,454	2.39%
Non-unique	739	1.21%
Unique	715	1.17%

H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Status	IssuedDate	Expire	Busin	Busin	Unit	UnitT	House	Street	City	Provi	Co	Posta	Local	Numb	FeePaid	Extra
Issued	2014-11-05 14:38	2015-12-31	Fitness Ce	Fitness Centre - Class 2	672	LEG-IN-BO	Vancouver BC	CA	V5Z 3Z4	Fairview	0					2019-07-
Issued	2014-11-13 12:16	2015-12-31	Wholesale Dealer	170	Unit	422	RICHARDS	Vancouver BC	CA	V6B 2Z4	Downtown	2				2019-07-
Issued	2014-11-14 09:42	2015-12-31	Retail Dea	Convenience	130	Unit	840	HOWE ST	Vancouver BC	CA	V6Z 2L2	Downtown	2			2019-07-
Issued	2014-11-14 14:50	2015-12-31	Contractor	Tiler				Burnaby BC	CA			2				2019-07-
Issued	2014-11-17 10:52	2015-12-31	Retail Dea	Furnishings			2025	W 4TH AV	Vancouver BC	CA	V6J 1N3	Kitsilano	3			2019-07-
Issued	2014-11-24 15:44	2015-12-31	Secondhand	Class 2 (Retail)			1192	BURRARD	Vancouver BC	CA	V6Z 1Y7	Downtown	7			2019-07-
Issued	2014-11-25 08:51	2015-12-31	Community Association				1170	BUTE ST	Vancouver BC	CA	V6E 1Z6	West End	10			2019-07-
Issued	2014-11-25 10:43	2015-12-31	Health Ser	Physician/	103	Unit	2025	W BROAD	Vancouver BC	CA	V6J 1Z6	Kitsilano	6			2019-07-
Issued	2014-11-27 16:19	2015-12-31	Contractor	Building					Vancouver BC	CA		Hastings-S	1			2019-07-
Issued	2014-12-01 16:29	2015-12-31	Secondary Suite - Permanent						Vancouver BC	CA		Victoria-Fr	000			2019-07-
Issued	2014-12-03 11:50	2015-12-31	Real Estate Dealer	205	Unit	2607	E 49TH AV	Vancouver BC	CA	V5S 1J9	Victoria-Fr	000				2019-07-
Issued	2014-12-08 10:02	2015-12-31	Retail Dea	Other					Vancouver BC	CA		Marpole	1			2019-07-
Issued	2014-12-08 10:42	2015-12-31	Fitness Ce	Fitness Ce	201	Unit	888	BURRARD	Vancouver BC	CA	V6Z 1X9	Downtown	10			2019-07-
Issued	2014-12-09 14:18	2015-12-31	Liquor Est	Class 2 66-150 Seats			2001	MACDON	Vancouver BC	CA	V6K 3Y2	Kitsilano	20			2019-07-
Issued	2014-12-10 10:13	2015-12-31	Landscape Gardener						Burnaby BC	CA			1			2019-07-
Issued	2014-12-10 12:23	2015-12-31	Office	Other					Vancouver BC	CA		Kitsilano	1			2019-07-
Issued	2014-12-11 11:39	2015-12-31	Hotel				1150	DENMAN	Vancouver BC	CA	V6G 2M9	West End	11			2019-07-
Issued	2014-12-12 13:18	2015-12-31	Secondary Suite - Permanent						Vancouver BC	CA		Renfrew-C	000			2019-07-
Issued	2014-12-12 13:21	2015-12-31	Employment Agency						Vancouver BC	CA		Kitsilano	2			2019-07-
Issued	2014-12-12 15:42	2015-12-31	Artist Live/Work Studio						Vancouver BC	CA		Strathcon	000			2019-07-
Issued	2014-12-15 15:15	2015-12-31	Contractor	Building			453	W 12TH A	Vancouver BC	CA	V5Y 1V4	Mount Ple	000			2019-07-
Issued	2014-12-17 10:39	2015-12-31	Retail Dea	Other	105	Unit	925	W GEORG	Vancouver BC	CA	V6C 3L2	Downtown	000			2019-07-
Issued	2014-12-19 11:07	2015-12-31	Community Associati		200	Unit	678	E HASTING	Vancouver BC	CA	V6A 1R1	Strathcon	26			2019-07-
Issued	2014-12-19 13:20	2015-12-31	Health Ser	Other					Vancouver BC	CA		West End	2			2019-07-

## SME Review of Initial Assessment

### LicenceRSN : Duplicates (High)

This is a serious issue with potential business impact. So this requires an immediate action. Will discuss this matter with I.T person and find out the cause and rectify this problem.

### Business Name : Null (Moderate)

This is a serious issue with potential business impact. So this requires an immediate action. Those Businesses should have one valid name these names were not available at the time of data entry as they were new in the market and have not registered at that point of time. Will discuss this matter with I.T person and find out the cause and rectify this problem. As you told none of these business's status is issued. So this concern don't need immediate attention.

### PostalCode: Invalid data Format (Moderate)

This is noted as an issue that has potential Impact on business, but does not need immediate attention as Postal Code attribute is rarely used in data modification, searching, insertion and deletion. Will follow a standard and rectify this issue in future.

### Status : Invalid Data (low)

This is a serious issue with potential business impact. As status plays a vital role in decision making. It's just 3 records and none of these businesses have licence status issued.

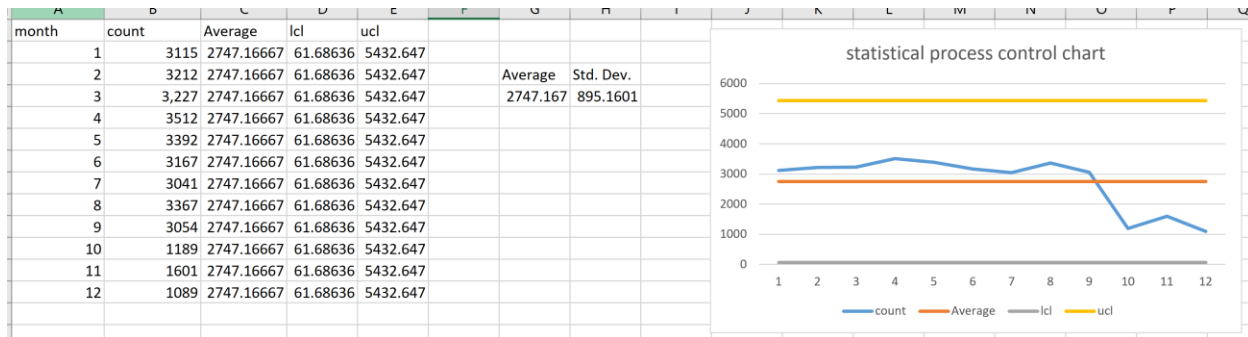
### Province : 2-Capitalized Letter Standard Not Followed (Moderate)

This is an issue of concern and will talk to I.T team to look into the matter and follow a standard naming convention.

### FeePaid: Null Values (moderate)

This is noted as an issue that has potential of business impact, and need immediate attention as can see from your screenshots that there are several businesses whose status is issued but fees paid is null.

## Further Research for the SME



## SME Review of Further Research

Classification	Issues falls in that particular category
Business Critical	LicenceRSN
Serious	BusinessName , PostalCode
Tolerable	FeePaid, Status
Acknowledged	ProvinceName

## SME Suggests Some DQ Rules

### LicenceRSN

LicenceRSN should be unique and not null

### Business Name

BusinessName must not be null if the record's Status=Issued and BusinessType<>Short-Term Rental.

### PostalCode

PostalCode must match the format of the Country for the record if the Status=Issued.

### Status

Status must exist in the StatusLookupTable.

### Province

Province must be two capitalized letters and must also exist in the ProvinceState table if Status=Issued.

### FeePaid

FeePaid must be greater than zero if the Status=Issued.