Project Report

New Car Buy Recommender System augmented with Data Mining

Masters of Technology in Intelligent Systems

Module: Reasoning Systems

Ajay Vikram Singh (A0020986B)
Rahul Jalan (A0195299H)

TABLE OF CONTENTS

Executive Summary

We, Singapore citizens and residents, are known globally for our 5 C desires- car, condo, cash, credit card, and country club. Car ownership in Singapore for Singaporean Residents is a cherished aspiration. [Ref 1] captures the sentiment in international press. Closer home, Singaporean love affair for car ownership is evident from weekend newspaper advertisements.

However, Car ownership is a costly affair for us, primarily because of the need to control traffic, ranking Singapore as one of the costliest places to own and maintain a car. [Ref 2]

Our "love" affair with car ownership is so skewed that MAS (monetary authority of Singapore) has to step in to enforce financial prudence in Singaporeans. [Ref 3]

A wide range of choices of car brands offering a bewildering range of possibilities across ride, comfort, safety, gadgetry etc are available to Singaporeans. This is thanks to our open and services driven economy.

Above all aspects, makes the buying decision process a very confusing and suboptimal affair, leaving much room for improvement.

Our project, New car buy recommendation system, is an effort to improve the car buy decisioning system in Singapore. Our recommendation system today suggests car budget and car make/type to look for, thus narrowing down the choice process quite significantly.

We used a combination of expert interviews as well as carefully designed surveys to define a set of rules that help make optimal decision on car buying. To capture the fuzziness in buying decision, we reinforced thus defined/mined rules with the concept of certainty factor.

We used the KIE for user input capture, KIE DROOLS for decision rule, and additionally Orange pack as well as Excel, Tableau and Google sheet analytics for offline data exploration and data mining.

In the reasoning systems project, we have now augmented the project (machine reasoning) to include rules defined by Data mining using decision trees

The solution we have presented is a first attempt to bring objectivity and optimality to car the decisioning process. Hence, we prepared a roadmap to capture further objectivity and augmented intelligence in car buying process.

Our achievement is happy Singapore car buyers, where they find the best ride suited to their needs and aspirations that the money they can realistically afford!

Problem

Definition

Which Car should I buy? Car Features, Japanese/European, fuel efficiency, second car, family needs are just some of the fuzzy questions that a car buyer goes through? What is the best decision considering various individual, family, societal needs, financial constraints Vs car ownership costs.

Above is the predicament household question in Singaporean car buyer minds. Unfortunately, no clear and robust decisioning system at hand.

Hence, we formulated the problem statement, thus:

A prospective car buyer weighs in various questions like above, and makes a subjective decision which we observe is not the most optimal.

Description

To further describe the problem involves an appreciation of following aspects for the problem space:

1. The buying process is subjective and hence fuzzy. The concept of Certainty Factor has to be introduced.
2. There is sufficient synthesized human knowledge on car buying that must be captured as part of decisioning system
3. Buying process has "crowd wisdom" and nuggets of knowledge have to be unearthed. Given that the buying process entails whole population of Singapore, hence surveys are clearly must and needed.

Scoping for the project

Inclusions

Business

1. Only Singapore market car buying behaviour and decisioning constraints are captured. Application formulation does not apply to other countries.
2. Singapore New car buying process is captured
3. Today's financial regulations on car loan
4. Expert interview and survey data mining

Technical

1. Application relies on making strong inferences from the survey and single expert interview (this is clearly an inferential limitation)
2. Application is offline today, based on KIE rules engine

3. Offline inclusion of survey data mining results

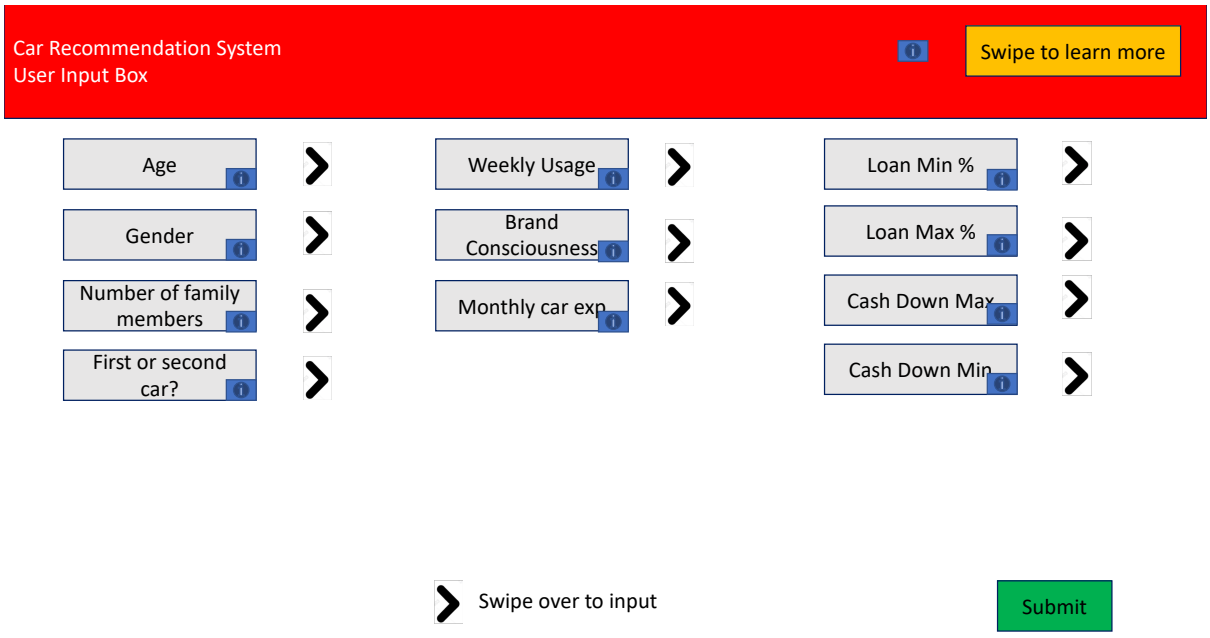
Exclusions

1. Not recommending individual model make (E.g. Toyota Camry 2.5 Hybrid Vs Mercedes Benz C-Class 180 Avantgarde). Models pricing and Model inventory/life cycle are volatile.
2. resale Car buying process
3. web interactive and public domain
4. Differentiation between car owners and prospective car buyers not made

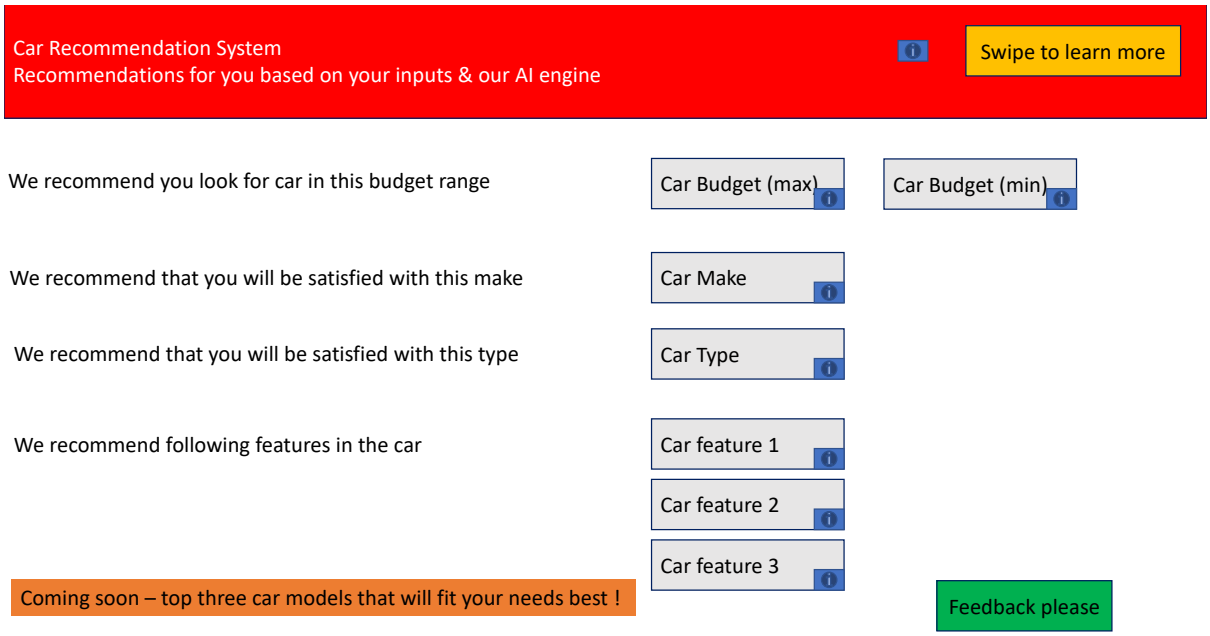
Please review the roadmap recommendations to address some of the exclusions listed above.

Inputs & Output dialogue screens (as delivered in Machine Reasoning)

User Input Form



Recommendation Screen



Reasoning system augmentation

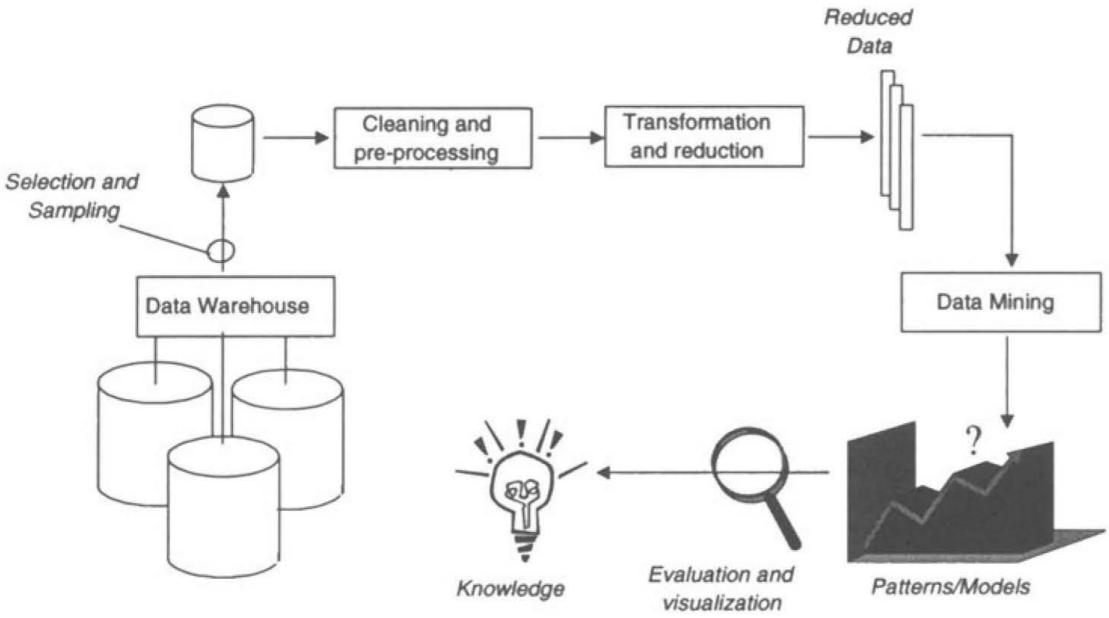
Reasoning system improvises "Car Make" and "Car Type" recommendations using rules identified by decision tree based data mining.

There are no changes in the input and the recommendation system screens.

Solution – Data Mining based augmentation

Knowledge Discovery Concepts

Data Mining and Knowledge Discovery in Databases (KDD) are highly evolving areas of research that are at the intersection of several disciplines, including statistics, databases, pattern recognition/AI, visualization, and high-performance and parallel computing. They promise to play an important role in the way people interact with databases, especially decision support databases where analysis and exploration operations are essential. KDD refers to the overall process of discovering useful knowledge from data while Data Mining is a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporating appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data.



Steps comprising the KDD process

Data Warehouse

Data Warehouses are central repositories of integrated data from one or more disparate sources. In an Enterprise setting they store current and historical data in one single place that are used for creating analytical reports for workers throughout the enterprise.

Selection and Sampling

Sampling is the selection of a subset of individuals from within a statistical population to estimate characteristics of the entire population. The advantages of sampling are lower cost and faster data collection than measuring the entire population.

Cleaning and Preprocessing

The phrase "garbage in, garbage out" is particularly applicable to data mining and ML projects. Data-gathering methods are often loosely controlled, resulting in some out-of-range values (e.g., Income: −1000), inappropriate data combinations (e.g., Sex: Male, Pregnant: Yes) missing values, etc. Data has to be carefully screened for such problems before data analysis takes place to avoid misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

Transformation and Reduction

Data reduction is the transformation of numerical or alphabetical digital information derived empirically or experimentally into a corrected, ordered, and simplified form. The basic concept is the reduction of multitudinous amounts of data down to the meaningful parts.

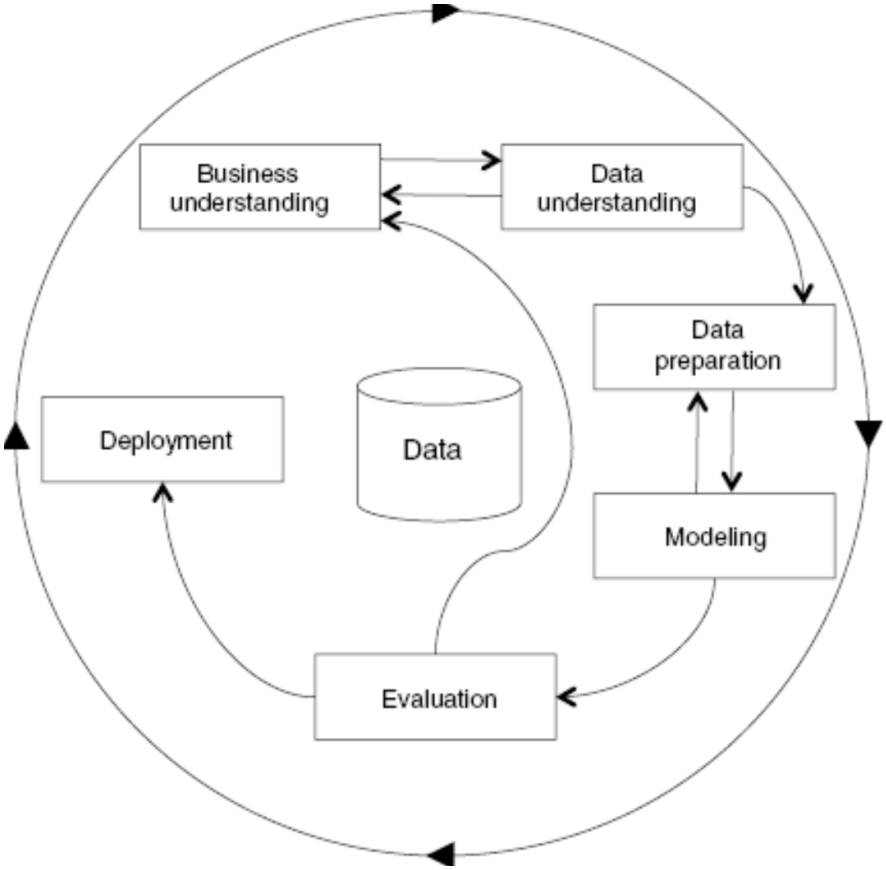
Data Mining

Data mining is a step in the KDD process consisting of applying computational techniques that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data. It is concerned with the algorithmic means by which patterns are extracted and enumerated from data. The overall KDD process includes the evaluation and possible interpretation of the "mined" patterns to determine which patterns may be considered new "knowledge."

The CRISP-DM Life cycle

There are various processes put forward to climb the data science pyramid. The most commonly used process is the Cross Industry Standard Process for Data Mining (CRISP-DM). In fact, the CRISP-DM has regularly been in the number-one spot in various industry surveys for a number of years. The primary advantage of CRISP-DM is that it is designed to be independent of any software, vendor, or data-analysis technique

The CRISP-DM life cycle consists of six stages: ***business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as shown in below figure***



The CRISP-DM life cycle

Business understanding and Data understanding

Here we try to define the goals of the project by understanding the business needs and the data that the business has available to it. This usually takes many iterations to ensure data appropriateness, and in an enterprise setting this involves multiple meetings with business-focused departments (e.g. – sales, marketing, operations) and with data-administrators.

Data Preparation

The focus of the data-preparation stage is the creation of a data set for analysis, and may involve integrating data sources from a number of databases. Once a data set has been created, the quality of the data needs to be checked and fixed. Outliers and missing values are typical data-quality problems. Checking the quality of the data is very important because errors in the data can have a serious effect on the performance of the data-analysis algorithms.

Modelling

At this stage automatic algorithms are used to extract useful patterns from the data and to create models that encode these patterns. Data scientists use a number of different ML algorithms to train a number of different models on the data set. In some cases an ML algorithm works by fitting a template model structure to a data set by setting the parameters of the template to good values for that data set (e.g., fitting a linear regression or neural network model to a data set). In other cases an ML algorithm builds a model in a piecewise fashion (e.g. growing a decision tree one node at a time beginning at the root node of the tree)

Evaluation and Deployment

These stages are focused on how the models fit the business and its processes. The tests run during the modeling stage are focused purely on the accuracy of the models for the data set. The *evaluation* phase involves assessing the models in the broader context defined by the business needs. Does a model meet the business objectives of the process? Is there any business reason why a model is inadequate? Based on the general assessment of the models, the main decision made during the evaluation phase is whether any of the models should be deployed in the business or another iteration of the CRISP-DM process is required to create adequate models. Assuming the evaluation process approves a model or models, the project moves into the final stage of the process: *deployment*. The deployment phase involves examining how to deploy the selected models into the business environment. This involves planning how to integrate the models into the organization's technical infrastructure and business processes.

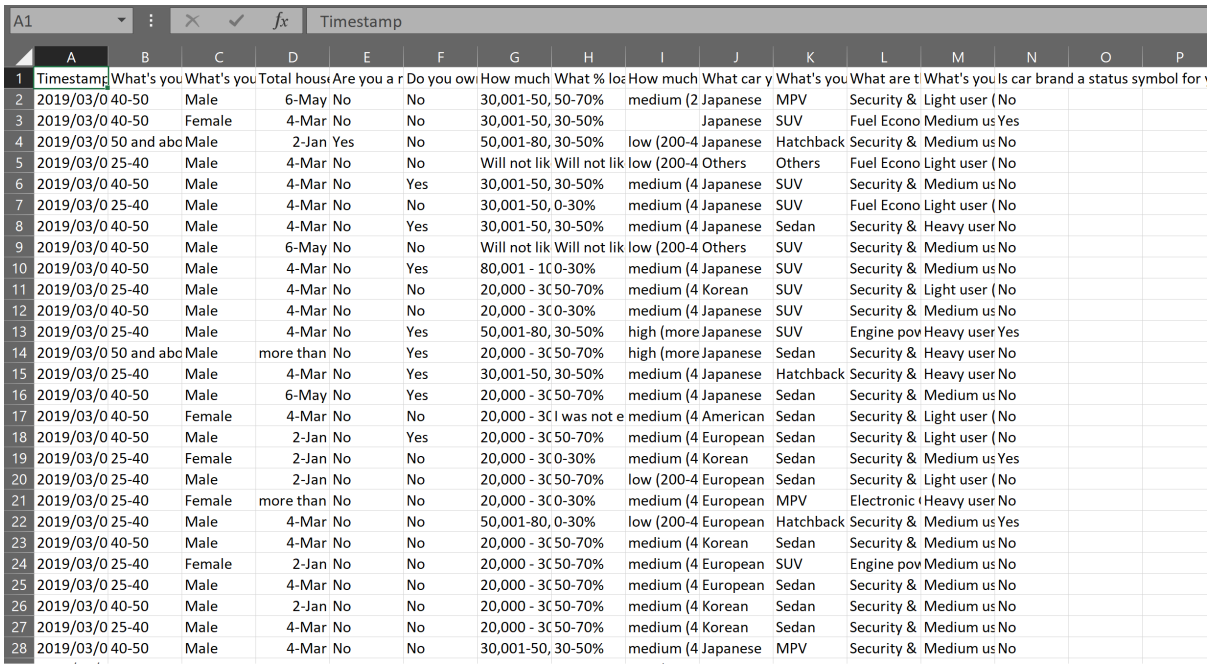
Its important to note that the whole process is iterative, After a project has developed and deployed a model, the model should be regularly reviewed to check that it still fits the business's needs and that it hasn't become obsolete

Application to Car Recommender System

We enhance the Car recommender system (CRS) that we developed using Machine Reasoning techniques by deriving the business rules by mining the survey data and using decision trees.

CRS Selection

In our case of CRS, we took the whole data set as our sample since the number of survey responders was within manageable limits (<60), hence we worked with whole dataset as a sample. The data was collected through an online survey which was open to public for a period of seven days. 59 people gave their inputs to a curated set of questions. Google Forms was used as tool which generates a .csv file with raw data from survey responses. The csv file is then loaded in excel for further steps.



A1 | Timestamp

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Timestamp	What's you	What's you	Total house	Are you a r	Do you ow	How much	What % loa	How much	What car y	What's you	What are t	What's you	Is car brand a status symbol for
2	2019/03/0	40-50	Male	6-May	No	No	30,001-50,	50-70%	medium (2	Japanese	MPV	Security &	Light user (	No
3	2019/03/0	40-50	Female	4-Mar	No	No	30,001-50,	30-50%		Japanese	SUV	Fuel Econo	Medium us	Yes
4	2019/03/0	50 and abo	Male	2-Jan	Yes	No	50,001-80,	30-50%	low (200-4	Japanese	Hatchback	Security &	Medium us	No
5	2019/03/0	25-40	Male	4-Mar	No	No	Will not lik	Will not lik	low (200-4	Others	Others	Fuel Econo	Light user (	No
6	2019/03/0	40-50	Male	4-Mar	No	Yes	30,001-50,	30-50%	medium (4	Japanese	SUV	Security &	Medium us	No
7	2019/03/0	25-40	Male	4-Mar	No	No	30,001-50,	0-30%	medium (4	Japanese	SUV	Fuel Econo	Light user (	No
8	2019/03/0	40-50	Male	4-Mar	No	Yes	30,001-50,	30-50%	medium (4	Japanese	Sedan	Security &	Heavy user	No
9	2019/03/0	40-50	Male	6-May	No	No	Will not lik	Will not lik	low (200-4	Others	SUV	Security &	Medium us	No
10	2019/03/0	40-50	Male	4-Mar	No	Yes	80,001 - 1(	0-30%	medium (4	Japanese	SUV	Security &	Medium us	No
11	2019/03/0	25-40	Male	4-Mar	No	No	20,000 - 3(	50-70%	medium (4	Korean	SUV	Security &	Light user (	No
12	2019/03/0	40-50	Male	4-Mar	No	No	20,000 - 3(	0-30%	medium (4	Japanese	SUV	Security &	Medium us	No
13	2019/03/0	25-40	Male	4-Mar	No	Yes	50,001-80,	30-50%	high (more	Japanese	SUV	Engine pov	Heavy user	Yes
14	2019/03/0	50 and abo	Male	more than	No	Yes	20,000 - 3(	50-70%	high (more	Japanese	Sedan	Security &	Heavy user	No
15	2019/03/0	25-40	Male	4-Mar	No	Yes	30,001-50,	30-50%	medium (4	Japanese	Hatchback	Security &	Heavy user	No
16	2019/03/0	40-50	Male	6-May	No	Yes	20,000 - 3(	50-70%	medium (4	Japanese	Sedan	Security &	Medium us	No
17	2019/03/0	40-50	Female	4-Mar	No	No	20,000 - 3(	I was not e	medium (4	American	Sedan	Security &	Light user (	No
18	2019/03/0	40-50	Male	2-Jan	No	Yes	20,000 - 3(	50-70%	medium (4	European	Sedan	Security &	Light user (	No
19	2019/03/0	25-40	Female	2-Jan	No	No	20,000 - 3(	0-30%	medium (4	Korean	Sedan	Security &	Medium us	Yes
20	2019/03/0	25-40	Male	2-Jan	No	No	20,000 - 3(	50-70%	low (200-4	European	Sedan	Security &	Light user (	No
21	2019/03/0	25-40	Female	more than	No	No	20,000 - 3(	0-30%	medium (4	European	MPV	Electronic (	Heavy user	No
22	2019/03/0	25-40	Male	4-Mar	No	No	50,001-80,	0-30%	low (200-4	European	Hatchback	Security &	Medium us	Yes
23	2019/03/0	40-50	Male	4-Mar	No	No	20,000 - 3(	50-70%	medium (4	Korean	Sedan	Security &	Medium us	No
24	2019/03/0	25-40	Female	2-Jan	No	No	20,000 - 3(	50-70%	medium (4	European	SUV	Engine pov	Medium us	No
25	2019/03/0	25-40	Male	4-Mar	No	No	20,000 - 3(	50-70%	medium (4	European	Sedan	Security &	Medium us	No
26	2019/03/0	40-50	Male	2-Jan	No	No	20,000 - 3(	50-70%	medium (4	Korean	Sedan	Security &	Medium us	No
27	2019/03/0	25-40	Male	4-Mar	No	No	20,000 - 3(	50-70%	medium (4	Korean	Sedan	Security &	Medium us	No
28	2019/03/0	40-50	Male	4-Mar	No	No	30,001-50,	30-50%	medium (4	Japanese	MPV	Security &	Medium us	No

CRS Preprocessing

In our CRS system, it involved formatting the data and deleting some incomplete entries. After preprocessing our data looked as below:

H19 | 50-70%

	A	B	C	D	E	F	G	H	I	J	K	
1	Survey duration	03.03.2019-09.03.2019										
2												
3												
4	Timestamp	What's your age group?	What's your gender?	Total household members (self, spouse, children, parents, maid, etc)	Are you a retiree?	Do you own a car	How much cash-down did you pay for your last car? (/or you will pay for your future car)	What % loan did you take on your last car buy? (/or you will pay for your future car)	How much are your other monthly car expenses? e.g. petrol, car tax, insurance (/or you will pay for your future car)	What car you own today? (/Or plan to own)	What's your car type? (/ or would be)	Wh (se
5	2019/03/03 8:08:14 PM GMT+8	40-50	Male	5-6	No	No	30,001-50,000 SGD	50-70%	medium (200-400 SGD)	Japanese	MPV	Sec
6	2019/03/03 8:10:27 PM GMT+8	40-50	Female	3-4	No	No	30,001-50,000 SGD	30-50%		Japanese	SUV	Fu
7	2019/03/03 8:11:14 PM GMT+8	50 and above	Male	1-2	Yes	No	50,001-80,000 SGD	30-50%	low (200-400 SGD)	Japanese	Hatchback	Sec
8	2019/03/03 8:37:22 PM GMT+8	25-40	Male	3-4	No	No	Will not like to reveal	Will not like to reve	low (200-400 SGD)	Others	Others	Fu
9	2019/03/03 8:38:46 PM GMT+8	40-50	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese	SUV	Sec
10	2019/03/03 8:39:32 PM GMT+8	25-40	Male	3-4	No	No	30,001-50,000 SGD	0-30%	medium (400-600 SGD)	Japanese	SUV	Fu
11	2019/03/03 8:41:09 PM GMT+8	40-50	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese	Sedan	Sec
12	2019/03/03 8:43:39 PM GMT+8	40-50	Male	5-6	No	No	Will not like to reveal	Will not like to reve	low (200-400 SGD)	Others	SUV	Sec
13	2019/03/03 8:45:46 PM GMT+8	40-50	Male	3-4	No	Yes	80,001 - 100,000 SGD	0-30%	medium (400-600 SGD)	Japanese	SUV	Sec
14	2019/03/03 8:46:56 PM GMT+8	25-40	Male	3-4	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Korean	SUV	Sec
15	2019/03/03 8:48:40 PM GMT+8	40-50	Male	3-4	No	No	20,000 - 30,000 SGD	0-30%	medium (400-600 SGD)	Japanese	SUV	Sec
16	2019/03/03 8:49:33 PM GMT+8	25-40	Male	3-4	No	Yes	50,001-80,000 SGD	30-50%	high (more then 400 SGD)	Japanese	SUV	En
17	2019/03/03 8:53:22 PM GMT+8	50 and above	Male	6+	No	Yes	20,000 - 30,000 SGD	50-70%	high (more then 400 SGD)	Japanese	Sedan	Sec
18	2019/03/03 8:55:04 PM GMT+8	25-40	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese	Hatchback	Sec
19	2019/03/03 8:56:35 PM GMT+8	40-50	Male	5-6	No	Yes	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Japanese	Sedan	Sec
20	2019/03/03 8:57:07 PM GMT+8	40-50	Female	3-4	No	No	20,000 - 30,000 SGD	I was not eligible for	medium (400-600 SGD)	American	Sedan	Sec
21	2019/03/03 9:01:40 PM GMT+8	40-50	Male	1-2	No	Yes	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	European	Sedan	Sec
22	2019/03/03 9:06:04 PM GMT+8	25-40	Female	1-2	No	No	20,000 - 30,000 SGD	0-30%	medium (400-600 SGD)	Korean	Sedan	Sec
23	2019/03/03 9:16:53 PM GMT+8	25-40	Male	1-2	No	No	20,000 - 30,000 SGD	50-70%	low (200-400 SGD)	European	Sedan	Sec
24	2019/03/03 9:27:17 PM GMT+8	25-40	Female	6+	No	No	20,000 - 30,000 SGD	0-30%	medium (400-600 SGD)	European	MPV	Ele
25	2019/03/03 9:34:43 PM GMT+8	25-40	Male	3-4	No	No	50,001-80,000 SGD	0-30%	low (200-400 SGD)	European	Hatchback	Sec
26	2019/03/03 9:37:25 PM GMT+8	40-50	Male	3-4	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Korean	Sedan	Sec
27	2019/03/03 9:51:02 PM GMT+8	25-40	Female	1-2	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	European	SUV	En
28	2019/03/03 9:57:35 PM GMT+8	25-40	Male	3-4	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	European	Sedan	Sec
29	2019/03/03 10:07:26 PM GMT+8	40-50	Male	1-2	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Korean	Sedan	Sec
30	2019/03/03 10:19:31 PM GMT+8	25-40	Male	3-4	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Korean	Sedan	Sec
31	2019/03/03 10:23:52 PM GMT+8	40-50	Male	3-4	No	No	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese	MPV	Sec
32	2019/03/03 11:08:29 PM GMT+8	25-40	Female	3-4	No	Yes	20,000 - 30,000 SGD	30-50%	high (more then 600 SGD)	European	SUV	Sec
33	2019/03/03 11:18:15 PM GMT+8	25-40	Female	1-2	No	Yes	Will not like to reveal	Will not like to reve	medium (400-600 SGD)	European	Hatchback	Sec
34	2019/03/04 8:05:46 AM GMT+8	25-40	Male	5-6	No	No	30,001-50,000 SGD	0-30%	high (more then 600 SGD)	Japanese	Sedan	Sec
35	2019/03/04 8:44:18 AM GMT+8	25-40	Male	3-4	No	Yes	30,001-50,000 SGD	50-70%	low (200-400 SGD)	Japanese	SUV	Sec

CRS Transformation

For CRS survey data, the transformation and reduction step involved converting various binary fields such as Retiree Info, car ownership etc into '0' and '1'. Also some additional column fields were designed to convert and capture interval type of data (e.g. – total household members) into columns with binary inputs, or cash down range into midpoint data.

Clipboard | Font | Alignment | Number | Styles

Q19 | Japanese

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3																	
4																	
5	#	Age	Gender	Total # / membe	1-2	3-4	5-6	6+	Retiree?	Own car	Cash down	Loan %	Cash Down (Estimated)	Loan % (Estimated)	Car value (Estimated)	Monthly Car Ex	Country
6	1	40-50	1	5-6	0	0	1	0	0	0	30,001-50,000 SGD	50-70%	40000	60%	$ 100,000	L	Japanese
7	2	40-50	0	3-4	0	1	0	0	0	0	30,001-50,000 SGD	30-50%	40000	40%	$ 66,667	M	Japanese
8	3	50 and above	1	1-2	1	0	0	0	1	0	50,001-80,000 SGD	30-50%	65000	40%	$ 108,333	L	Japanese
9	4	25-40	1	3-4	0	1	0	0	0	0	Will not like to reveal	Will not like to reve	65000	30%	$ 92,857	L	Others
10	5	40-50	1	3-4	0	1	0	0	0	1	30,001-50,000 SGD	30-50%	40000	40%	$ 66,667	M	Japanese
11	6	25-40	1	3-4	0	1	0	0	0	0	30,001-50,000 SGD	0-30%	40000	30%	$ 57,143	M	Japanese
12	7	40-50	1	3-4	0	1	0	0	0	1	30,001-50,000 SGD	30-50%	40000	40%	$ 66,667	M	Japanese
13	8	40-50	1	5-6	0	0	1	0	0	0	Will not like to reveal	Will not like to reve	65000	30%	$ 92,857	L	Others
14	9	40-50	1	3-4	0	1	0	0	0	1	80,001 - 100,000 SGD	0-30%	90000	30%	$ 128,571	M	Japanese
15	10	25-40	1	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Korean
16	11	40-50	1	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	0-30%	25000	30%	$ 35,714	M	Japanese
17	12	25-40	1	3-4	0	1	0	0	0	1	50,001-80,000 SGD	30-50%	65000	40%	$ 108,333	H	Japanese
18	13	50 and above	1	6+	0	0	0	1	0	1	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	H	Japanese
19	14	25-40	1	3-4	0	1	0	0	0	1	30,001-50,000 SGD	30-50%	40000	40%	$ 66,667	M	Japanese
20	15	40-50	1	5-6	0	0	1	0	0	1	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Japanese
21	16	40-50	0	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	I was not eligible fo	25000	0%	$ 25,000	M	America
22	17	40-50	1	1-2	1	0	0	0	0	1	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Europea
23	18	25-40	0	1-2	1	0	0	0	0	0	20,000 - 30,000 SGD	0-30%	25000	30%	$ 35,714	M	Korean
24	19	25-40	1	1-2	1	0	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	L	Europea
25	20	25-40	0	6+	0	0	0	1	0	0	20,000 - 30,000 SGD	0-30%	25000	30%	$ 35,714	M	Europea
26	21	25-40	1	3-4	0	1	0	0	0	0	50,001-80,000 SGD	0-30%	65000	30%	$ 92,857	L	Europea
27	22	40-50	1	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Korean
28	23	25-40	0	1-2	1	0	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Europea
29	24	25-40	1	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Europea
30	25	40-50	1	1-2	1	0	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Korean
31	26	25-40	1	3-4	0	1	0	0	0	0	20,000 - 30,000 SGD	50-70%	25000	60%	$ 62,500	M	Korean
32	27	40-50	1	3-4	0	1	0	0	0	0	30,001-50,000 SGD	30-50%	40000	40%	$ 66,667	M	Japanese
33	28	25-40	0	3-4	0	1	0	0	0	1	20,000 - 30,000 SGD	30-50%	25000	40%	$ 41,667	H	Europea

CRS Data Mining

We reviewed multiple data mining algorithms considering in mind the data and data model at hand.

We selected decision tree as it allows for clear rules definition and is easily explainable. This allows for further roadmap definition and accuracy improvement initiatives to combine various *knowledge discovery and rules augmentation.*

The predictors chosen are the following:

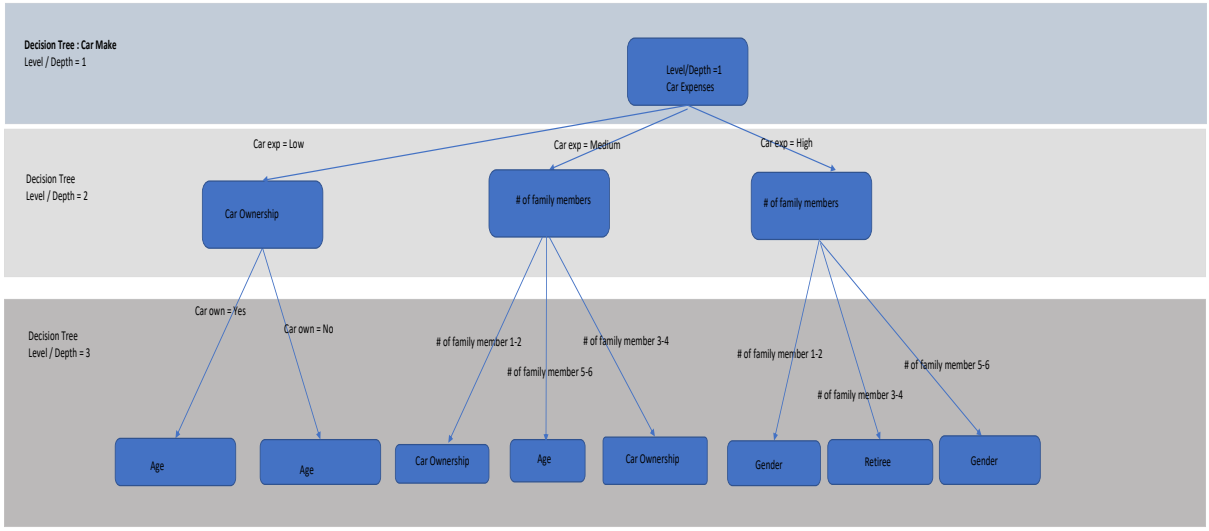
1. Age : Age bracket of the person. Here we have converted to age brackets, and subsequent hot encoding to make the input discrete variables. One hot encoding ranges are [25-40, 40-50, 50+].
2. Gender: Male or Female. Binary variable
3. # of family members: Total number of family members. Here we again converted to discrete inputs using one hot encoding. Ranges are [1-2, 3-4,5-6,6+]. One hot encoding.
4. Retiree: Binary variable
5. Car ownership: Binary variable
6. Car Expenses: Range: Low, Medium, High. Type one hot encoding.

We used ID3 algorithm that defines the decision nodes based on improving the information gain.

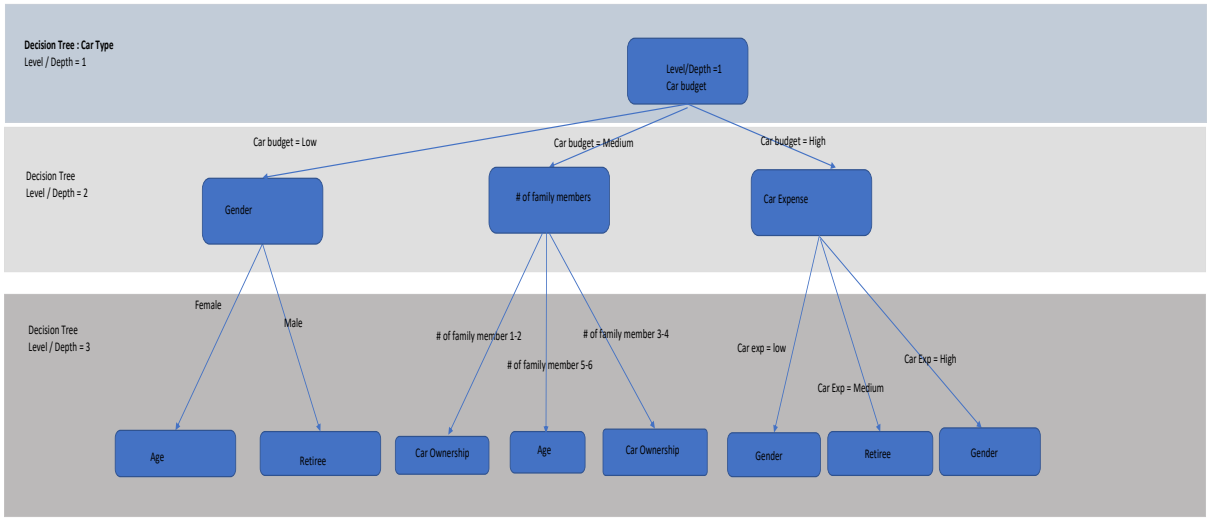
While we have ran the decision tree based on limited data samples, we are conscious that the decision tree **should avoid overlearning**. Hence the decision tree, has been restricted to 3 levels.

Here is the finalized decision trees:

Decision tree for “Car Make” Predicted variable/target:



Decision tree for "Car Type" Predicted variable/target:



CRS Interpretation/Evaluation

The rules identified through data mining help validate and augment the following:

a) Human expert based rules definition
b) Confidence factor based rules definition

Through decision tree, which is an explainable data mining model, we gain additional confidence in rules definition and their commercial rollout.

Project Critical Analysis

Key Survey and Expert Findings

- Housing type does not matter in car selection as was based on expert feedback
- Gadgets and Music does not matter much to the survey population. We believe that it's due to a reasonably OK music system in most of the cars in market, and is not a buying factor anymore.
- Women overwhelmingly look for 'fuel economy' of car
- Men overwhelmingly look for 'safety and security' in the car.
- Retirees overwhelmingly look for Japanese car
- The application does not consider the working status because it was explicitly suggested by expert that it does not influence car buying behaviour. However, detailed data based studies can be further taken to explore alternate hypothesis.
- Most of the buyer base is 40-50 years bracket
- Retirees mostly want to buy car with full cash down payment
- Young overwhelmingly take maximum loans

Functionality and system short comings, pending issues

- Distinction of European cars – wide range of quality difference between European car brands
- Distinction between sedan categories – ideally luxury and mass market sedans are different markets
- specific tastes like Racing, Vintage cars
- Resale value of the car was not captured in the survey. While heuristics is available, i.e. Toyota and Mercedes fair well on resales value in Singapore market, since there was no supporting data from survey we have excluded it from user inputs. However, based on user inputs, we factor based on calibration.
- Next best alternative (NBA), top car make, type suggestions not implemented. E.g. Toyota Camry 2.5 Hybrid is the top recommendation based on inputs and knowledge decisioning system (i.e. Car recommendation system)
- Weekly/Daily fluctuations due to COE, flash sales
- OPC (off-peak car) or full COE based recommendations
- Ability to give meaningful recommendation with users providing only partial information
- System handles only male and female preferences (Transgender and other social /family structure preferences not curated)
- For retirees, we have not considered loan guarantors for loan. This is based on expert advise as well that most (~100%+) do cash down buying.

System Limitations

- Lacks Visual appeal due to using KIE forms.
- User interactivity limited to capability afforded by KIE forms.
- System can't go live online (public internet domain) due to KIE capabilities and Public domain/public server hosting and budgets
- Inputs that shall have a finite range to represent, hence choices are made binary (e.g. security preference that can range from low, medium, high or on a scale of 1-10)

Future Project Enhancements

Scope Enhancements

- Singapore Used car market inclusion
- Expanding recommendations from car make, type, features, to top matching 3 cars (NBA) available in the Singapore market at that given point in time
- Expanding to nearby countries, e.g. South East Asian markets

Functional Enhancements

- Representative surveys for Singapore car buying target population
- Machine learning module augmentation to identify
- Additional experts feedback
- Hypothesis testing scenarios

Technical Enhancements

- Launching the mobile interface and web server (AWS) with a scale up on demand backend i.e. AWS IaaS/PaaS.
- Machine Learning module for offline integration based on model re-runs (surveys and user feedbacks) with AWS sagemaker.
- Integration with car dealers (e.g. C&C, Borneo motors etc) and car information aggregation websites like sgcarmart.com

References

[Ref 1]: Singaporean love for cars
http://money.com/money/4716340/worlds-most-expensive-city-cost-of-living-car-ownership/

[Ref 2]: Singapore is consistently the costliest place to buy car
https://www.startrescue.co.uk/news/motoring-news-2/10-most-expensive-countries-to-buy-and-run-a-car
https://www.sgcarmart.com/news/article.php?AID=10371

[Ref 3]:
Singapore government regulations on car loans

http://www.mas.gov.sg/News-and-Publications/Media-Releases/2016/MAS-Eases-Rules-on-Motor-Vehicle-Financing.aspx

http://www.ifaq.gov.sg/MAS/TOPICS/BANKING/Vehicle_Loans/7419#FAQ_176191

[Ref 4]:
Expert interview: Jay, Singaporean Car Sales manager with more than a decade of successful car selling experience in Singapore. Audio transcripts submitted.

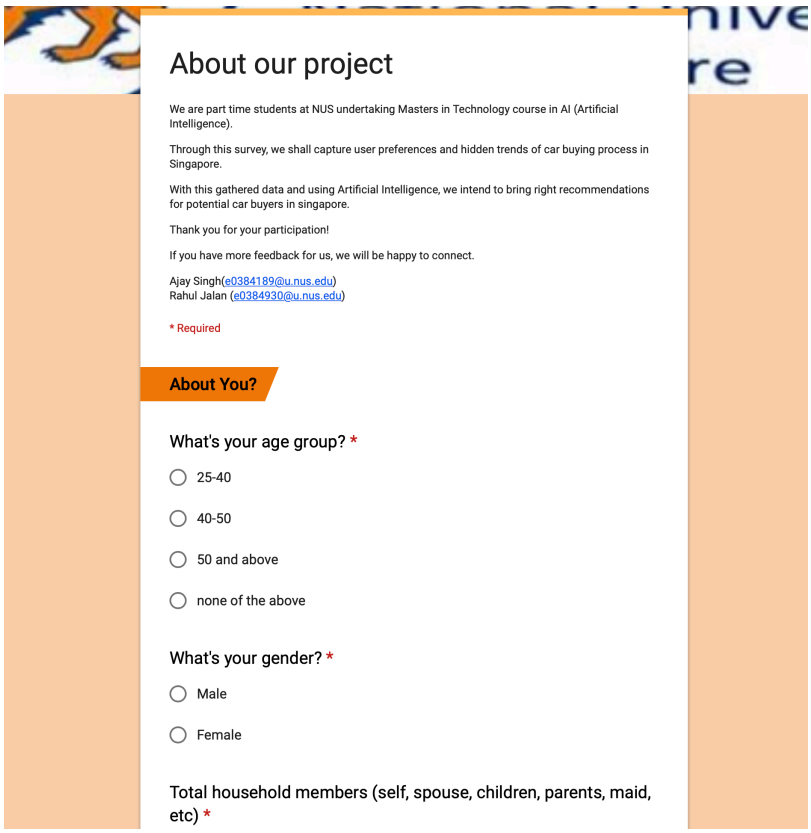
[Ref 5]:
SGCarMart.com: is the leading Singapore car information portal.

[Ref 6]:
Survey to collect end buyer behaviour
https://goo.gl/forms/Frt1Rh3q9ImSXKkY2

Appendices

Appendix A: Survey

Survey Form:



About our project

We are part time students at NUS undertaking Masters in Technology course in AI (Artificial Intelligence).

Through this survey, we shall capture user preferences and hidden trends of car buying process in Singapore.

With this gathered data and using Artificial Intelligence, we intend to bring right recommendations for potential car buyers in singapore.

Thank you for your participation!

If you have more feedback for us, we will be happy to connect.

Ajay Singh(e0384189@u.nus.edu)
Rahul Jalan (e0384930@u.nus.edu)

* Required

About You?

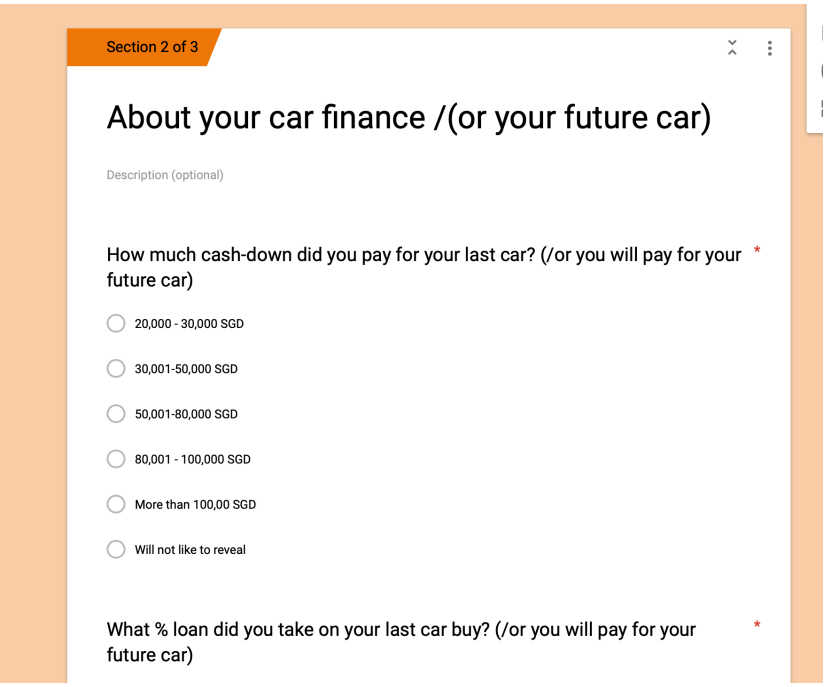
What's your age group? *

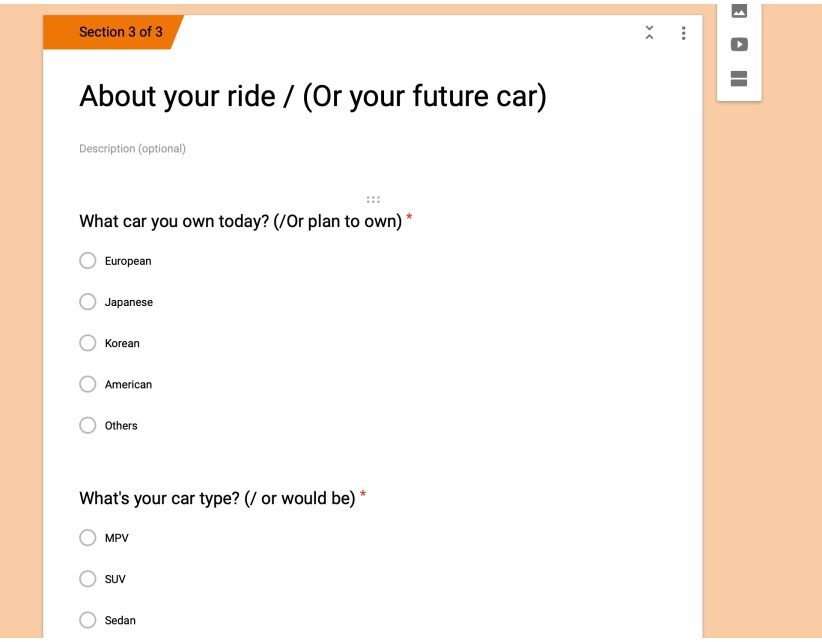
- 25-40
- 40-50
- 50 and above
- none of the above

What's your gender? *

- Male
- Female

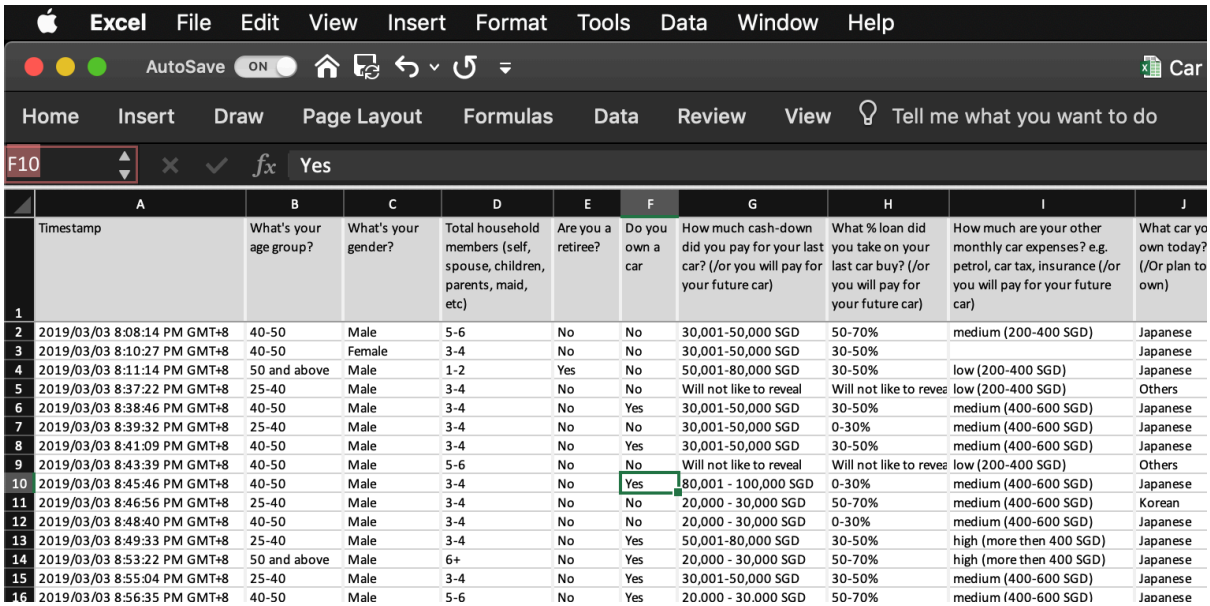
Total household members (self, spouse, children, parents, maid, etc) *





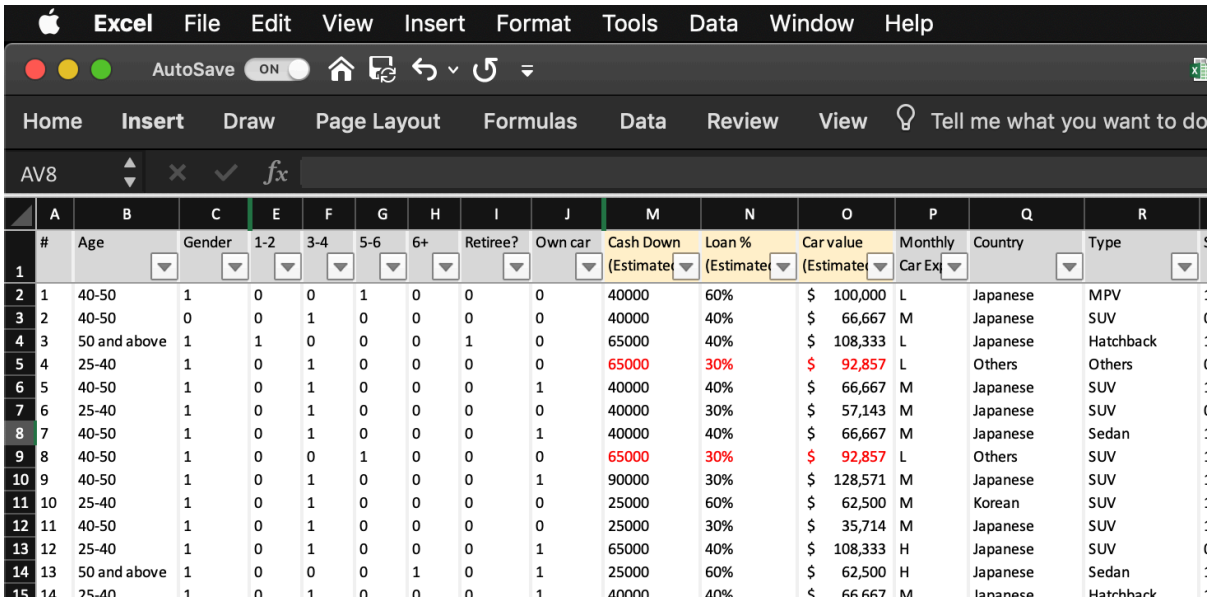
Survey raw data:

Details attached in project zip folder



	A	B	C	D	E	F	G	H	I	J
1	Timestamp	What's your age group?	What's your gender?	Total household members (self, spouse, children, parents, maid, etc)	Are you a retiree?	Do you own a car	How much cash-down did you pay for your last car? (/or you will pay for your future car)	What % loan did you take on your last car buy? (/or you will pay for your future car)	How much are your other monthly car expenses? e.g. petrol, car tax, insurance (/or you will pay for your future car)	What car yo own today? (/Or plan to own)
2	2019/03/03 8:08:14 PM GMT+8	40-50	Male	5-6	No	No	30,001-50,000 SGD	50-70%	medium (200-400 SGD)	Japanese
3	2019/03/03 8:10:27 PM GMT+8	40-50	Female	3-4	No	No	30,001-50,000 SGD	30-50%		Japanese
4	2019/03/03 8:11:14 PM GMT+8	50 and above	Male	1-2	Yes	No	50,001-80,000 SGD	30-50%	low (200-400 SGD)	Japanese
5	2019/03/03 8:37:22 PM GMT+8	25-40	Male	3-4	No	No	Will not like to reveal	Will not like to revea	low (200-400 SGD)	Others
6	2019/03/03 8:38:46 PM GMT+8	40-50	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese
7	2019/03/03 8:39:32 PM GMT+8	25-40	Male	3-4	No	No	30,001-50,000 SGD	0-30%	medium (400-600 SGD)	Japanese
8	2019/03/03 8:41:09 PM GMT+8	40-50	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese
9	2019/03/03 8:43:39 PM GMT+8	40-50	Male	5-6	No	No	Will not like to reveal	Will not like to revea	low (200-400 SGD)	Others
10	2019/03/03 8:45:46 PM GMT+8	40-50	Male	3-4	No	Yes	80,001 - 100,000 SGD	0-30%	medium (400-600 SGD)	Japanese
11	2019/03/03 8:46:56 PM GMT+8	25-40	Male	3-4	No	No	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Korean
12	2019/03/03 8:48:40 PM GMT+8	40-50	Male	3-4	No	No	20,000 - 30,000 SGD	0-30%	medium (400-600 SGD)	Japanese
13	2019/03/03 8:49:33 PM GMT+8	25-40	Male	3-4	No	Yes	50,001-80,000 SGD	30-50%	high (more then 400 SGD)	Japanese
14	2019/03/03 8:53:22 PM GMT+8	50 and above	Male	6+	No	Yes	20,000 - 30,000 SGD	50-70%	high (more then 400 SGD)	Japanese
15	2019/03/03 8:55:04 PM GMT+8	25-40	Male	3-4	No	Yes	30,001-50,000 SGD	30-50%	medium (400-600 SGD)	Japanese
16	2019/03/03 8:56:35 PM GMT+8	40-50	Male	5-6	No	Yes	20,000 - 30,000 SGD	50-70%	medium (400-600 SGD)	Japanese

Survey cleaned data:



	A	B	C	E	F	G	H	I	J	M	N	O	P	Q	R
1	#	Age	Gender	1-2	3-4	5-6	6+	Retiree?	Own car	Cash Down (Estimate	Loan % (Estimate	Car value (Estimate	Monthly Car Ex	Country	Type
2	1	40-50	1	0	0	1	0	0	0	40000	60%	$ 100,000	L	Japanese	MPV
3	2	40-50	0	0	1	0	0	0	0	40000	40%	$ 66,667	M	Japanese	SUV
4	3	50 and above	1	1	0	0	0	1	0	65000	40%	$ 108,333	L	Japanese	Hatchback
5	4	25-40	1	0	1	0	0	0	0	65000	30%	$ 92,857	L	Others	Others
6	5	40-50	1	0	1	0	0	0	1	40000	40%	$ 66,667	M	Japanese	SUV
7	6	25-40	1	0	1	0	0	0	0	40000	30%	$ 57,143	M	Japanese	SUV
8	7	40-50	1	0	1	0	0	0	1	40000	40%	$ 66,667	M	Japanese	Sedan
9	8	40-50	1	0	0	1	0	0	0	65000	30%	$ 92,857	L	Others	SUV
10	9	40-50	1	0	1	0	0	0	1	90000	30%	$ 128,571	M	Japanese	SUV
11	10	25-40	1	0	1	0	0	0	0	25000	60%	$ 62,500	M	Korean	SUV
12	11	40-50	1	0	1	0	0	0	0	25000	30%	$ 35,714	M	Japanese	SUV
13	12	25-40	1	0	1	0	0	0	1	65000	40%	$ 108,333	H	Japanese	SUV
14	13	50 and above	1	0	0	0	1	0	1	25000	60%	$ 62,500	H	Japanese	Sedan

Details attached in project zip folder

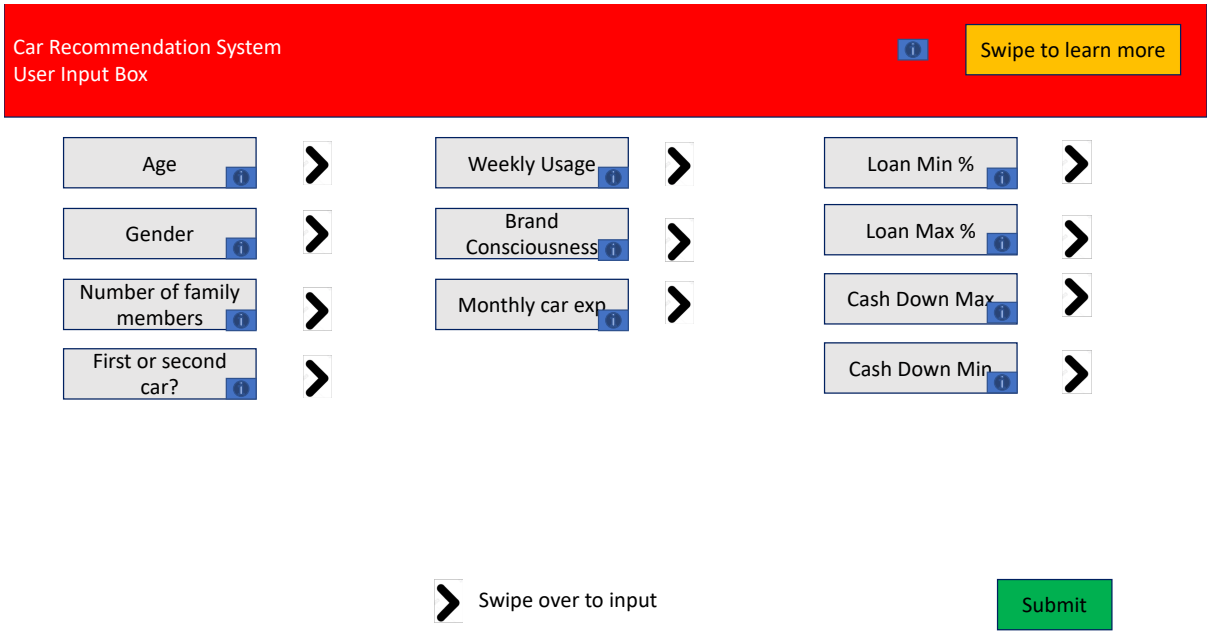
Survey Google analytics report

Details attached in project zip folder

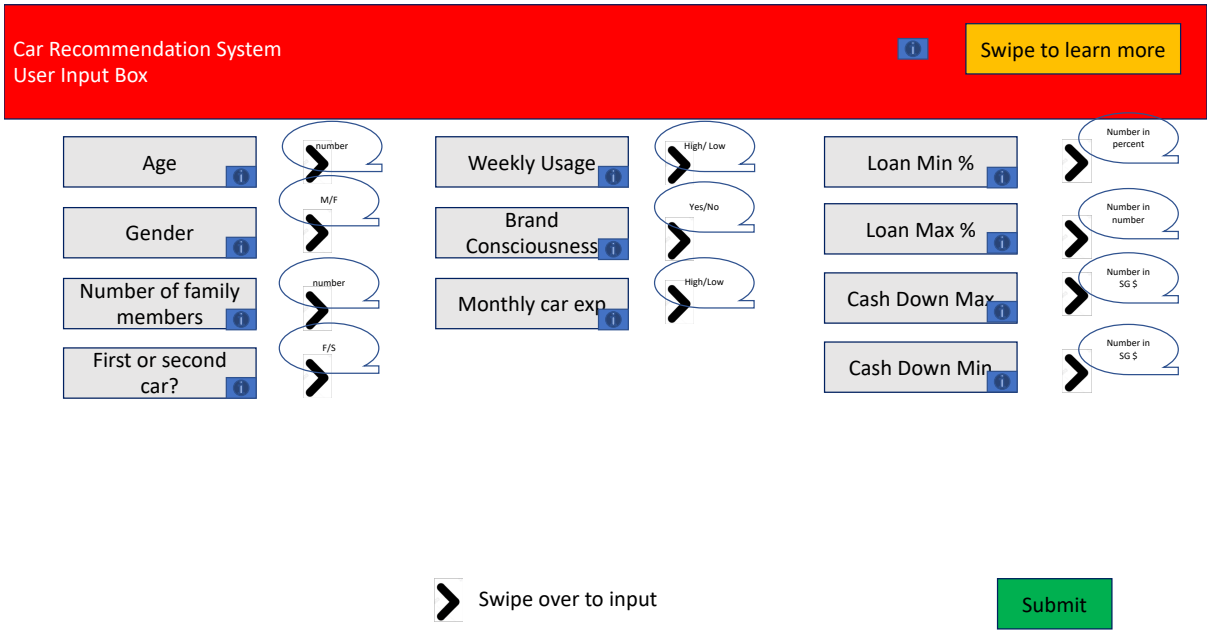
Appendix B: Car Recommendation System User Guide

User is guide to a dialogue box upon open the application on mobile or through a public URL As discussed previously, the MVP is planned for KIE only. Functionality has to be exposed through a web API.

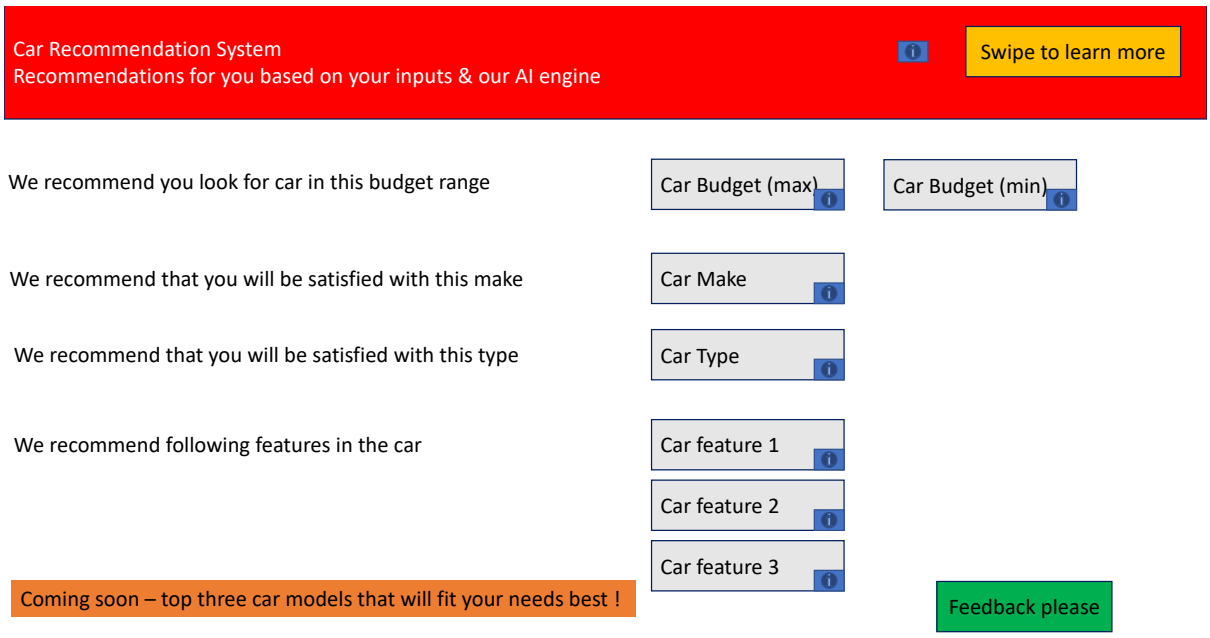
The input screen will appear like below:



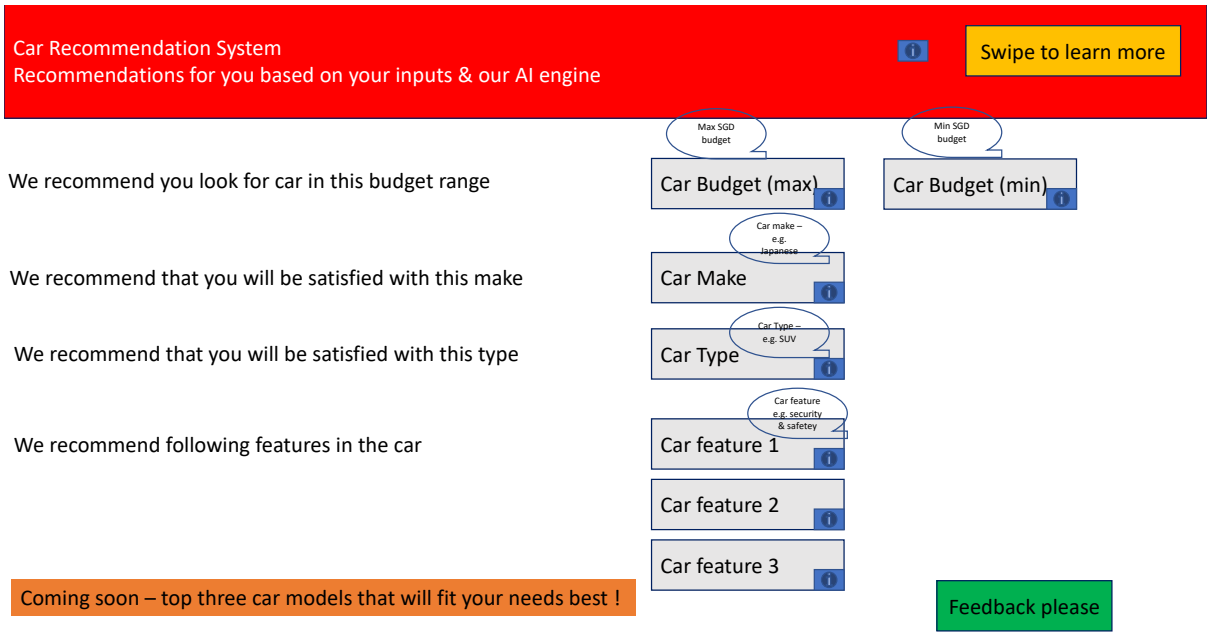
At the input screen, please input all the input fields by swiping on the forward arrow mark. The inputs are either true/false or based on some numeric range, like highlighted below!



The output, in today's MVP, will appear like this.



Point towards the information icon will provide additional details about a recommendation item.


Car Recommendation System
Recommendations for you based on your inputs & our AI engine
Swipe to learn more
We recommend you look for car in this budget range
Max SGD budget
Car Budget (max)
Min SGD budget
Car Budget (min)
We recommend that you will be satisfied with this make
Car make – e.g. Japanese
Car Make
We recommend that you will be satisfied with this type
Car Type – e.g. SUV
Car Type
We recommend following features in the car
Car feature e.g. security & safetey
Car feature 1
Car feature 2
Car feature 3
Coming soon – top three car models that will fit your needs best !
Feedback please

Appendix C: Sample Scenario

Scenario: Value buyer, aged 30, male

A car buyer with following requirements captured in the user input form, ref appendix B for the user's guide, for user input dialogue box in mobile app/browser.

Input	Description	Data type	User inputs
Age	Age of buyer	numeric	30
Gender	Gender of buyer	Binary	Male
Num_People_family	Number of family members in the hous	Numeric	2
is_second_car	is the car being bought second car	Binary	No
Cash_min	Min cash user thinks has for car buy	Numeric	25000
Cash_max	Max cash user think has for car buy	Numeric	30000
Loan_min	Min loan user think shall need for car bu	percent	50
Loan_max	Max loan user think shall need for car b	percent	80
Weekly_Usage	Weekly car usage	Binary	1
Brand_Consciousness	Is brand consciousness a factor in decisi	Binary	1
monthly_car_expenses	Monthly expense include fuel, maintena	Binary	0

Results in decision process calculation, ref Table 7, flowchart for the rule definition engine to calculate CF

Input	Description	Data type	User inputs	Basis CF calculations
Age	Age of buyer	numeric	30	
Gender	Gender of buyer	Binary	Male	CF_SS == 1
Num_People_family	Number of family members in the hous	Numeric	2	CF_Type_SUV/CF_Type_MPV == -1
is_second_car	is the car being bought second car	Binary	No	not relevant for CF influence in this case
Cash_min	Min cash user thinks has for car buy	Numeric	25000	$83,333
Cash_max	Max cash user think has for car buy	Numeric	30000	$100,000
Loan_min	Min loan user think shall need for car bu	percent	50	CF_30_loan == -0.5
Loan_max	Max loan user think shall need for car b	percent	80	CF_70_loan == 1
Weekly_Usage	Weekly car usage	Binary	1	CF_features_economy == 1
Brand_Consciousness	Is brand consciousness a factor in decisi	Binary	1	CF_European == 1, CF_Jap == 0.8
monthly_car_expenses	Monthly expense include fuel, maintena	Binary	0	CF_European (updated) == -1, CF_Jap == 0.8

The car recommendation system, as defined in Appendix B, user guide will make following recommendations. With these recommendations, the budget is narrowed down, the car make, type and features narrowed down as well.

Ouput	Output bracket	Possible values	Recommendations
Price	Price Min	Minimum car pric	$83,333
	Price Max	Maximum car pric	$100,000
Car country make	Country of car make	Japenese	Japanese
		European	No
		Korean	Does not matter
		American	No
		Malaysian	No
Car Type	Type of car	Sedan	Yes
		SUV	No
		MPV	No
		Hatchback	Yes
		Stationwagon	
Car Features	Recommended car features	Safety and securit	Yes
		Fuel Economy	Yes
		Engine	Does not matter

Once we link the app to current market models and pricing, we can zero down to best 2-3 options available (future enhancement captured in Section Future Project Enhancements)

Appendix D: Glossary

MAS : Monetary Authority of Singapore
LTA : Land Transport Authority of Singapore
COE : certificate of entitlement. Granted by LTA Singapore. Each car must carry the COE. COE is a major component of car price.
TCO : Total cost of ownership
EMI : Equal Monthly instalment
CF : Certainty factor
BHP : Brake Horse Power
MVP : Minimum Viable Product
KIE : Knowledge is everything
DDoS : Distributed Denial of Service
C&C : Cycle and Carriage
NBA : next best alternatives
AWS : Amazon Web Services
IaaS : Infrastructure as a Service
PaaS : Platform as a Service
SUV : Sports Utility Vehicle
MPV : Multi Purpose Vehicle
CRS : Car Recommendation System

Appendix E. Solution (as delivered in Machine Reasoning Module)

The solution involves a three-step knowledge definition and synthesis process:

1. Acquisition & Discovery
2. Structuring
3. Modelling & Representation

Knowledge Discovery & Acquisition

Knowledge Acquisition constituted following 3 step process:

a) Step 1: Self-study and self-articulation: There are two areas of knowledge acquisition here:
   a. Individual and family/Friends buying experience and post facto (post buying challenges). This formed the dimensions of decision making process, the dimensions nature – whether definite or fuzzy contributors to buying logic. This step also involved reviewing government rules and regulations regarding loan eligibility [Ref 3] and minimum cash down for car buying.
   b. Leading car portals: [Ref 5]. We leveraged SGCarmart portal and paid special attention to their car comparison and car catalogue pages. These give curated car buyer knowledge acquisition process

b) Step 2: Expert feedback: Equipped with our understanding and appreciation of car buying process, its dimensions and challenges, we sought feedback of a Subject Matter Expert. [Ref 4]. We conducted an in-depth interview conducted with Jay, who carries a sound decade plus continued experience in Singapore car buying market as a senior new car sales manager. Synthesizing the interview outcomes, we:

   a. Confirmed evident rules in step 1
   b. un-earthed un-conventional buying rules, amounting to 100% certainty factor. We also identified "grey" buying areas in our interview. Besides, we got a good understanding on finances and finance behaviour related to car buying process. E.g. women buyers predominantly look for fuel-efficient cars.

Row Labels	Sum of Fuel	Count of Gender	
Female	15	16	94%
Male	30	42	
(blank)			
Grand Total	45	58	

Table 1: Female drivers a CF_features_economy = 1
(see table 3 for the whole list of CFs)

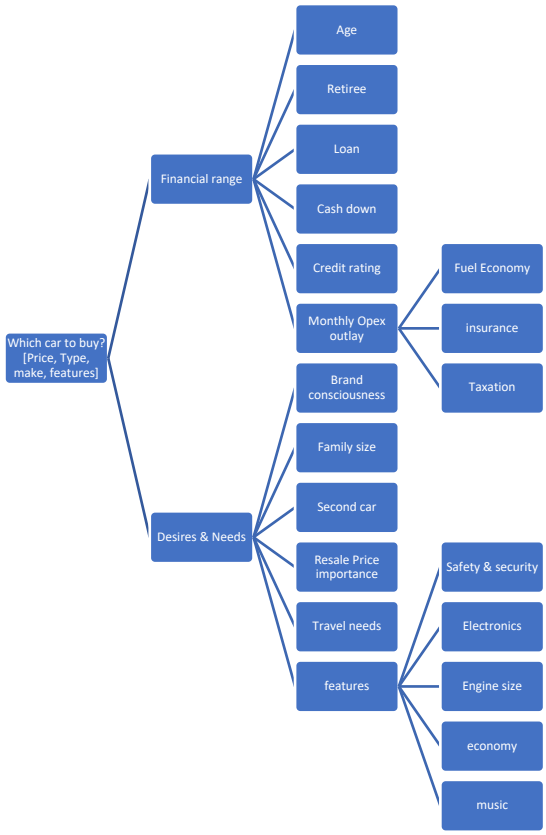
c) Step 3: End-user Survey: Having understood key matters related to car buying process, and also knowing grey areas in our thus acquired knowledge above, we went to survey creation. Sample Survey questionnaire is in [Ref 6]

In Summary, our project's knowledge acquisition tree looks like below:

(Figure 1: Knowledge Acquisition for car buying recommendation system)

Knowledge Structuring – Dependency diagram

Synthesized Knowledge from Knowledge discovery and acquisition process is represented by dependency diagram (figure 2) below.



(Figure 2: Dependency diagram: how to buy a new car in Singapore)

Knowledge Modelling & Representation

Buying budget behaviour and calculation: Derived Rules

Car budget recommendation is based on knowledge rules extracted from 3 step processes of Knowledge discovery and acquisition. Here following facts are important for defining the recommendation on car buying budget:

- Singapore regulation stipulates a maximum 70% loan. Min 30% must be cash down
- Financial lending companies and car companies ecosystem brings more incentives if a minimum 30% percentage of car value is taken as loan
- Younger people generally prefer maximum loan
- Older people generally prefer no to minimum loan

Accordingly, we defined, confidence factor (CF) to model end buyer fuzziness in financial clarity Vs the observed behaviour in field and government rules. The CF defined are:

1. CF to take 70% loan = CF_70_loan
2. CF to take 30% loan = CF_30_loan

Based on user inputs, CFs are updated based on a combination of survey and expert rules.

For the three identified age brackets, the mappings are as follows:

Scenario	Customer age group mapping	CF_70_loan	CF_30_loan
Scenario 1	Age > 40 & Age < 50	0	0
Scenario 2	Age > 50	-1	0.8
Scenario 3	Age < 40	1	-0.5

Table 2: Mapping of Age to CF 70/30

Associated car budget logic for each of the scenarios:

Scenario 1:

IF

(CF_70_loan == 0 && CF_30_loan == 0)

THEN

Price_Max = Cash_Max/(1- loan_max)

Price_Min = Cash_Min/(1-loan_min)

Where Max/Min represents user inputs/gut feel on max min loan and cash down

Scenario 2:

IF

(CF_70_loan <= -0.5 && CF_30_loan >= 0.5)

THEN

Price_Max = Cash_Max

Price_Min = Cash_Min/(1-30%)

Scenario 3:

IF

(CF_70_loan >= 0.5 && CF_30_loan <= -0.5)

THEN

Price_Max = Cash_Max/(1- 70%)

Price_Min = Cash_Min/(1-70%)

Certainty factors

As discussed in earlier sections, the buying process is inherently fuzzy and is inclusive of multiple factors. Hence, we have defined the following variables to capture the fuzziness.

Certainty Factor (CF)	Explanation (leaning towards)	min	max	initialized to
CF_30_loan	taking 30% loan	-1	1	0
CF_70_loan	taking 70% loan	-1	1	0
CF_Make_EUR	a European make car - e.g. Mercedes	-1	1	0
CF_Make_JP	a Japanese make car - e.g. toyota	-1	1	0
CF_Make_MY	a malayisa make car	-1	1	0
CF_Make_KOR	a korean make car	-1	1	0
CF_Make_American	an American make car	-1	1	0
CF_Type_Hatch	a hatchbatch e.g. Honda Jazz	-1	1	0
CF_Type_SUV	an SUV e.g. Subaru forester	-1	1	0
CF_Type_MPV	an MPV like Honda Stream	-1	1	0
CF_Type_Sedan	a Sedan e.g. Toyota camry	-1	1	0
CF_Type_Stationwagon	a stationwagon e.g. Volvo V60	-1	1	0
CF_features_SS	safety and security leading car	-1	1	0
CF_features_economy	higher fuel economy car	-1	1	0
CF_features_engine	higher engine performance i.e. high Torque,	-1	1	0
CF_features_gadgets	electronic gadgets like vision, car parking aid	-1	1	0
CF_features_music	higher quality music instruments	-1	1	0

Table 3: Certainty factor for the car recommender system

User inputs:

Car recommendation system end user inputs are captured as follows:

Input	Description	Data type	Initiatized to	Values
Age	Age of buyer	numeric	18	18-80
Gender	Gender of buyer	Binary	0	M = 1, F= 0
Num_People_family	Number of family members in the hous	Numeric	0	
is_second_car	is the car being bought second car	Binary	0	0= 1st car, 1= 2nd car
Cash_min	Min cash user thinks has for car buy	Numeric	0	0 to 500000
Cash_max	Max cash user think has for car buy	Numeric	0	0 to 500000
Loan_min	Min loan user think shall need for car bu	percent	0%	0%-70%
Loan_max	Max loan user think shall need for car b	percent	70%	0%-70%
Weekly_Usage	Weekly car usage	Binary	0	0 = low usage, 1 = high usage
Brand_Consciousness	Is brand consciousness a factor in decisi	Binary	0	1=brand conscious, 0 = not
monthly_car_expenses	Monthly expense include fuel, maintena	Binary	0	1= high, 0 = low

Table 4: User inputs table for the car recommender system

Car recommendation system application outputs:

Car recommendation system outputs are captured by following table:

Defined Rules

Based on Expert interview and Survey results (Certainty Factor), below diagram's capture the evolution of the CFs based on inputs (table 4) given by an end user of the car recommendation system application.

First, a few examples of user survey derived Certainty factors:

Example 1: user survey derived CF recommendation

Sum of Fuel	Column Labels							
Row Labels	American		European	Japanese	Korean	Others	(blank)	Grand Total	
25-40			8	7	4	5		24	
40-50		1	2	9	3	2		17	
50 and above			1	3				4	75%
(blank)									
Grand Total		**1**	**11**	**19**	**7**	**7**		**45**	

Table 5: >50 years old users car make preference

Example 2: user survey derived CF recommendation

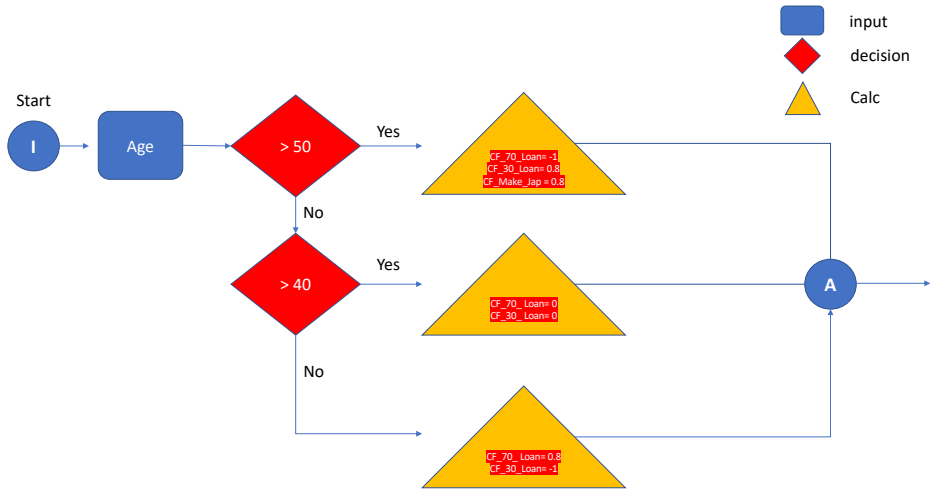
Older people prefer Japanese, resulting in derived rule CF_Make_Jap = 0.8 (see detailed flowchart in tab

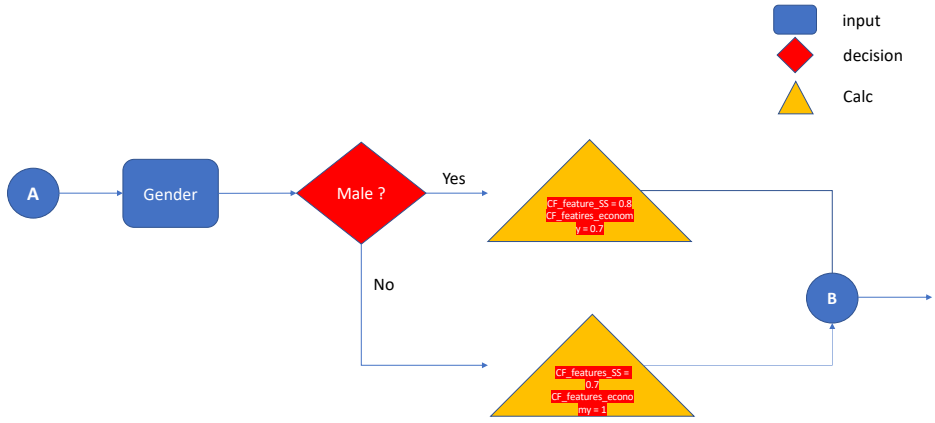
Row Labels	Sum of SS	Count of Gender	
Female	11	16	
Male	36	42	86%
(blank)			
Grand Total	**47**	**58**	

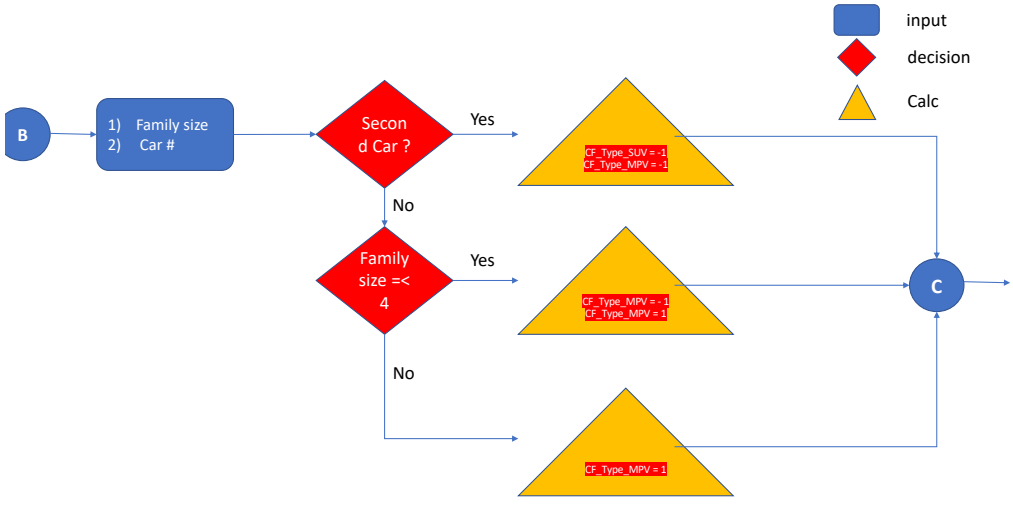
Table 6: Males prefer safety and security.

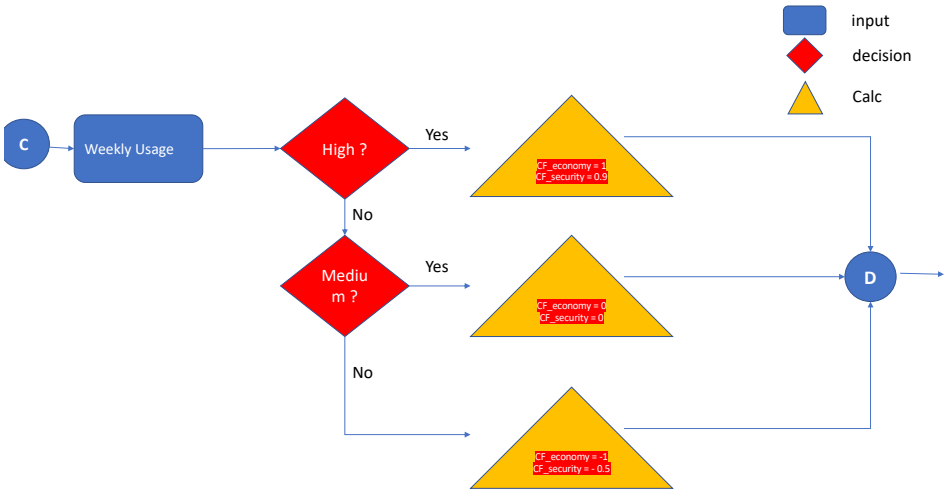
Refer table 1 for female preference

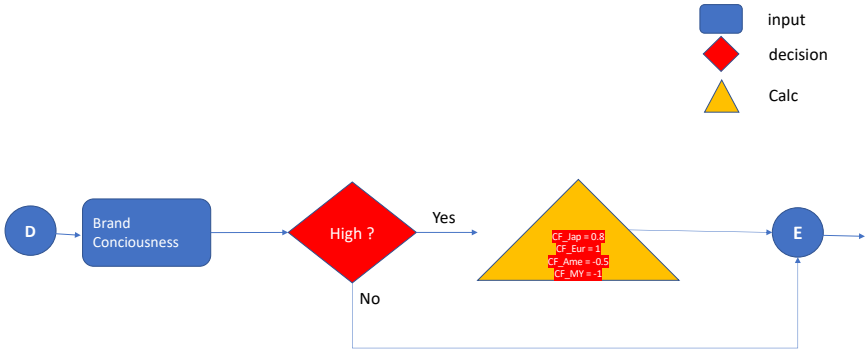
Finally, here is the flowchart for user input based CF influence.


input
decision
Calc
Start
I
Age
> 50
Yes
CF_70_Loan= -1
CF_30_Loan= 0.8
CF_Make_Jap = 0.8
No
> 40
Yes
CF_70_ Loan= 0
CF_30_ Loan= 0
A
No
CF_70_ Loan= 0.8
CF_30_Loan= -1


input
decision
Calc
A
Gender
Male ?
Yes
CF_feature_SS = 0.8
CF_featires_economy = 0.7
No
B
CF_features_SS = 0.7
CF_features_economy = 1


input
decision
Calc
B
1) Family size
2) Car #
Second Car ?
Yes
CF_Type_SUV = -1
CF_Type_MPV = -1
No
Family size =< 4
Yes
CF_Type_MPV = -1
CF_Type_MPV = 1
C
No
CF_Type_MPV = 1


input
decision
Calc
C
Weekly Usage
High ?
Yes
CF_economy = 1
CF_security = 0.9
No
Mediu
m ?
Yes
CF_economy = 0
CF_security = 0
D
No
CF_economy = -1
CF_security = - 0.5


input
decision
Calc
D
Brand
Conciousness
High ?
Yes
CF_Jap = 0.8
CF_Eur = 1
CF_Ame = -0.5
CF_MY = -1
E
No

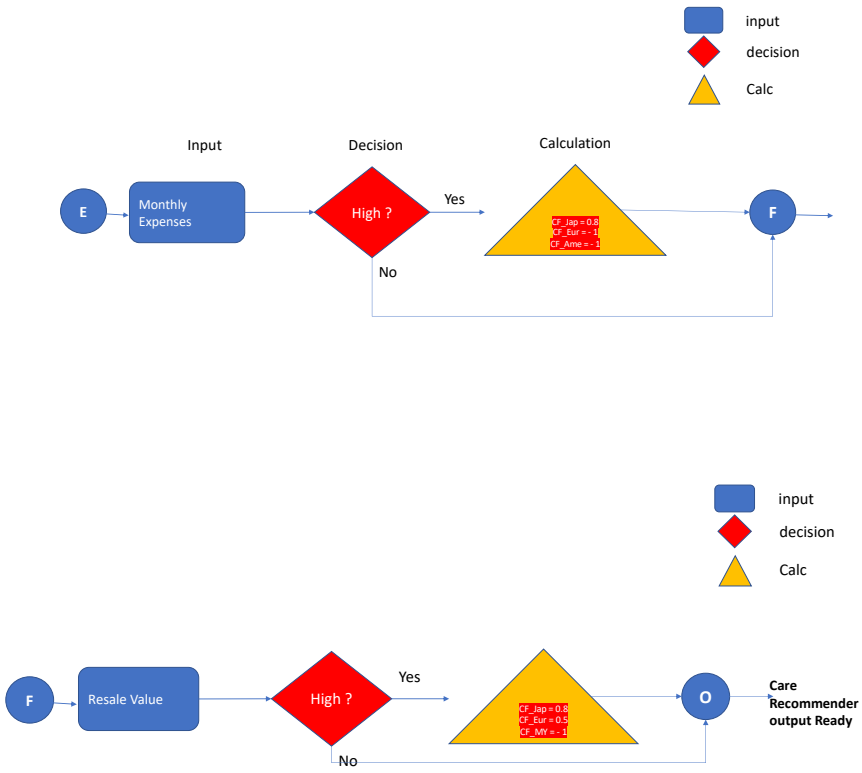


Table 7: Flowchart for Rule Definition engine to update the Certainty Factor

Finalized CFs will then generate the recommendation to the car recommendation system application end user.

Car recommendation system Outputs

Ouput	Output bracket	Possible values	Comments
Price	Price Min	Minimum car pric	car price bracket
	Price Max	Maximum car pric	car price bracket
Car country make	Country of car make	Japenese	In order of priority
		European	in order of priority
		Korean	in order of priority
		American	in order of priority
		Malaysian	in order of priority
Car Type	Type of car	Sedan	In order of priority
		SUV	In order of priority
		MPV	In order of priority
		Hatchback	In order of priority
		Stationwagon	In order of priority
Car Features	Recommended car features	Safety and securit	yes no
		Fuel Economy	yes no
		Engine	Yes no

Table 8: Output setup for the car recommendation system (in current version)

System Design:

In today's MVP stage, the system logic is defined in KIE. This includes definition of the

- User input form (as per the user input table defined in Table 4)
- Task definitions and Data Objects
- Process flow including input corrections/data correction
- And finally guided decision tables to provide the end recommendations (as per table 8 for car recommendation system output)

System Architecture considerations:

There are no system architecture considerations as we are in MVP stage. For the application to be launched to end users, following developments have to be defined:

- Mobile and web interface for user interactivity (Input and Output)
- Web server hosting (for the MVP functionality and logic)
- Security provisions for unauthorized access, DDoS etc attacks

Appendix F - Project Critical Analysis (as in Machine Reasoning Solution)

Key Survey and Expert Findings

- Housing type does not matter in car selection as was based on expert feedback
- Gadgets and Music does not matter much to the survey population. We believe that it's due to a reasonably OK music system in most of the cars in market, and is not a buying factor anymore.
- Women overwhelmingly look for 'fuel economy' of car
- Men overwhelmingly look for 'safety and security' in the car.
- Retirees overwhelmingly look for Japanese car
- The application does not consider the working status because it was explicitly suggested by expert that it does not influence car buying behaviour. However, detailed data based studies can be further taken to explore alternate hypothesis.
- Most of the buyer base is 40-50 years bracket
- Retirees mostly want to buy car with full cash down payment
- Young overwhelmingly take maximum loans

Functionality and system short comings, pending issues

- Distinction of European cars – wide range of quality difference between European car brands
- Distinction between sedan categories – ideally luxury and mass market sedans are different markets
- specific tastes like Racing, Vintage cars
- Resale value of the car was not captured in the survey. While heuristics is available, i.e. Toyota and Mercedes fair well on resales value in Singapore market, since there was no supporting data from survey we have excluded it from user inputs. However, based on user inputs, we factor based on calibration.
- Next best alternative (NBA), top car make, type suggestions not implemented. E.g. Toyota Camry 2.5 Hybrid is the top recommendation based on inputs and knowledge decisioning system (i.e. Car recommendation system)
- Weekly/Daily fluctuations due to COE, flash sales
- OPC (off-peak car) or full COE based recommendations
- Ability to give meaningful recommendation with users providing only partial information
- System handles only male and female preferences (Transgender and other social /family structure preferences not curated)
- For retirees, we have not considered loan guarantors for loan. This is based on expert advise as well that most (~100%+) do cash down buying.

System Limitations

- Lacks Visual appeal due to using KIE forms.
- User interactivity limited to capability afforded by KIE forms.

- System can't go live online (public internet domain) due to KIE capabilities and Public domain/public server hosting and budgets
- Inputs that shall have a finite range to represent, hence choices are made binary (e.g. security preference that can range from low, medium, high or on a scale of 1-10)

Appendix G - Future Project Enhancements (as in Machine Reasoning Solution)

Scope Enhancements

- Singapore Used car market inclusion
- Expanding recommendations from car make, type, features, to top matching 3 cars (NBA) available in the Singapore market at that given point in time
- Expanding to nearby countries, e.g. South East Asian markets

Functional Enhancements

- Representative surveys for Singapore car buying target population
- Machine learning module augmentation to identify
- Additional experts feedback
- Hypothesis testing scenarios

Technical Enhancements

- Launching the mobile interface and web server (AWS) with a scale up on demand backend i.e. AWS IaaS/PaaS.
- Machine Learning module for offline integration based on model re-runs (surveys and user feedbacks) with AWS sagemaker.
- Integration with car dealers (e.g. C&C, Borneo motors etc) and car information aggregation websites like sgcarmart.com