

**ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS**

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
DEPARTMENT OF MANAGEMENT, SCIENCE & TECHNOLOGY
MSc BUSINESS ANALYTICS

“Data Warehousing & Visualization”

Full Name: STAMATIOS SIDERIS

Register Number: f2822113

&

Full Name: ORESTIS LOUKOPOULOS

Register Number: f2822104

ATHENS, 2021

Table of Contents

1.	Business Case	page 3
2.	Report's Targets	page 3
3.	Data Source	page 3
4.	Data Cleaning	page 4
5.	Data Import	page 8
6.	Data Cube Deployment	page 13
7.	Business Analysis	page 16

Business Case

Auto+ is an upcoming auto auction-sales company which is planning to open its first warehouse/dealership, in the USA, at the beginning of 2016. To minimize the market entry risk, Auto+ asked our consulting company, OS Intelligence S.A., to provide them with a market analysis report of the car auctions market, in specific states of the USA, for the first half of 2015, as well as with visualization tools for better understanding and future use by the company.

Report's Targets

To provide Auto+ with useful insights about the preferences of the average used car auction sales customer in the USA as well as, suggest top car models to be supplied and best location to get their warehouse/dealership registered in, based on these preferences.

Data Source

The dataset is called “used-car-auction-prices” and is provided in the following URL: <https://www.kaggle.com/tunguz/used-car-auction-prices>, by www.kaggle.com. The dataset includes historical data for used car auction sales in the USA for three months in 2014 & seven months in 2015.

The dataset has the following 16 columns:

- year : The construction year of the car
- make: The manufacturer of the car
- model: The model of the car
- trim: The trim of the car
- body: The body type of the car
- transmission: Whether the car is automatic or manual
- vin: A unique code for the car
- state: The state where the car was sold
- condition: The condition of the car
- odometer: The total number of miles the car has completed
- color: The exterior color of the car
- interior: The interior color of the car
- seller: The name of the seller
- mmr: The Manheim Market Report (MMR) is the premier indicator of wholesale price
- sellingprice: The price that the car was sold
- saledate: The date when the car was sold

We used R-studio to have a quick insight of our data set. Using the summary function, we analyzed each column to find the variable type, minimum/maximum and mean values. Moreover, we observed each column for missing values.

Summary Function

```
> summary(car[, -c(18:20)])
  autoID      constructionY      make      model      trim
Min.   :    1      Min.   :1982  Length:558837  Length:558837  Length:558837
1st Qu.:139710  1st Qu.:2007    Class :character  Class :character  Class :character
Median :279419  Median :2012    Mode  :character  Mode  :character  Mode  :character
Mean   :279419  Mean   :2010
3rd Qu.:419128  3rd Qu.:2013
Max.   :558837  Max.   :2015

  body      transmission      vin      state      condition
Length:558837  Length:558837  Length:558837  Length:558837  Min.   :1.000
Class :character  Class :character  Class :character  Class :character  1st Qu.:2.700
Mode  :character  Mode  :character  Mode  :character  Mode  :character  Median :3.600
                                          Mean   :3.425
                                          3rd Qu.:4.200
                                          Max.   :5.000
                                          NA's   :11794

  odometer      color      interior      seller      mmr
Min.   :    1      Length:558837  Length:558837  Length:558837  Min.   :    25
1st Qu.: 28371    Class :character  Class :character  Class :character  1st Qu.:   7100
Median : 52254    Mode  :character  Mode  :character  Mode  :character  Median : 12250
Mean   : 68320
3rd Qu.: 99109
Max.   :999999
NA's   :94

  sellingprice      saledate      sales_time
Min.   :    1      Length:558837  Length:558837
1st Qu.: 69000    Class :character  Class :character
Median :12100    Mode  :character  Mode  :character
Mean   :13611
3rd Qu.:18200
Max.   :230000
```

Missing Values Check

```
> car$constructionY[car$constructionY==""] <- NA ;sum(is.na(car$constructionY))
[1] 0
> car$make[car$make==""] <- NA ;sum(is.na(car$make))
[1] 498
> car$model[car$model==""] <- NA ;sum(is.na(car$model))
[1] 517
> car$trim[car$trim==""] <- NA ;sum(is.na(car$trim))
[1] 777
> car$body[car$body==""] <- NA ;sum(is.na(car$body))
[1] 2925
> car$transmission[car$transmission==""] <- NA ;sum(is.na(car$transmission))
[1] 12544
> car$vin[car$vin==""] <- NA ;sum(is.na(car$vin))
[1] 0
> car$state[car$state==""] <- NA ;sum(is.na(car$state))
[1] 0
> car$color[car$color==""] <- NA ;sum(is.na(car$color))
[1] 3802
> car$interior[car$interior==""] <- NA ;sum(is.na(car$interior))
[1] 7545
> car$seller[car$seller==""] <- NA ;sum(is.na(car$seller))
[1] 0
> car$saledate[car$saledate==""] <- NA ;sum(is.na(car$saledate))
[1] 0
> sum(is.na(car))
[1] 40496
```

The dataset includes 12 character-variables and 6 numeric-variables. Highest number of missing values is observed in the “transmission” variable with 12544 empty cells. We changed the label of column “year” to “construction” and we also added a new column “autoID” which indicates each unique sale.

Data Cleaning

We used Excel to clean our data. First, we observed that there were mismatched values in columns trim and body. Hence, we fixed this misplacement by transferring the mismatched values to the correct column and move all the remaining columns one column to the left.

[illegible]

AutoID Column Addition

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	id	year	make	model	trim	body	transmission	vin	state	condit	odome	color	interior	seller
2	1	2015	Kia	Sorento	LT	SUV	automatic	5xvt6a69f5666472	CA	5	16639	white	black	kia motors america, inc
3	2	2015	Kia	Sorento	UX	SUV	automatic	5xvt6a69f5661319	CA	5	9938	white	beige	kia motors america, inc
4	3	2015	BMW	3 Series	328i SULEV	Sedan	automatic	wa3ba31c0a41163531	CA	4	1331	gray	black	financial services reman
5	4	2015	Volvo	S60	T5	Sedan	automatic	vv1612baf1310087	CA	4	1182	white	black	volvo na rep/world omni
6	5	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba9b2c57e41297531	CA	4	2641	gray	black	financial services reman
7	6	2015	Nissan	Altima	2.5 S	Sedan	automatic	1nfal3apfr3260313	CA	1	5554	gray	black	financial services reman
8	7	2014	BMW	M5	Base	Sedan	automatic	wba9f5c1e59593089	CA	3	1943	black	black	the hertz corporation
9	8	2014	Chevrolet	Cruze	1LT	Sedan	automatic	1g1pc5sb272128460	CA	3	2801	black	enterprise vehicle excha	
10	9	2015	Audi	A6	2.5	Sedan	automatic	wvauff3f3030345	CA	4	9557	white	black	audition vjeo
11	10	2014	Chevrolet	Camaro	LT	Convertible	automatic	2t1fbf3d579e218789	CA	3	4807	red	black	d/m auto sales inc
12	11	2014	Audi	A6	3.0T Prestige quattro	Sedan	automatic	wvaufgac0602916	CA	4	1444	black	black	desert auto trade
13	12	2015	Kia	Optima	SE	Sedan	automatic	svgnw6a37g353538	CA	4	2044	red	gray	financial finance
14	13	2015	Ford	Fusion	SE	Sedan	automatic	3f6p9dphd4154753	CA	2	3559	white	beige	enterprise vehicle excha
15	14	2015	Kia	Sorento	UX	SUV	automatic	5xvt6a69f5661407	CA	5	14634	silver	black	kia motors america, inc
16	15	2014	Chevrolet	Cruze	1LT	Sedan	automatic	1g1pc5sb272120091	CA	3	1568	black	black	financial services reman
17	16	2015	Nissan	Altima	2.5 S	Sedan	automatic	1nfal3apfr3260313	CA	2	1398	black	black	enterprise vehicle excha
18	17	2014	Hyundai	Sonata	2.4	Sedan	automatic	5npe2af4f0001562	CA	4	8311	red	gray	avix tra
19	18	2014	Audi	Audi	2.0T Premium Plus quattro	Sedan	automatic	wa1ffawc085074	CA	3	7981	white	black	avix tra north scottsdale
20	19	2014	BMW	Camaro	LS	Coupe	automatic	5npe2af4f0001562	CA	3	7981	white	black	avix tra
21	20	2014	BMW	6 Series	650i	Coupe	automatic	5xvt6a39e9314494	CA	1	1341	black	black	wells fargo dealer serv
22	21	2014	BMW	3 Series	328i	Convertible	automatic	wba9p9b53ed160260	CA	3	8810	black	black	the hertz corporation
23	22	2015	BMW	5 Series	528i	Sedan	automatic	wa1b53p309103921	CA	3	1538	silver	black	enterprise vehicle excha
24	23	2015	BMW	5 Series	528i	Sedan	automatic	wba9s4c51e051631	CA	2	2950	black	black	financial services reman
25	24	2015	Chevrolet	Equinox	LT	Convertible	automatic	2t1fbf3d51e9134662	CA	3	39450	black	black	financial services reman
26	25	2014	Audi	A3	1.8 TFSI Premium	Sedan	automatic	wvaugff7f1002327	CA	4	5826	gray	black	audition north scottsdale
27	26	2014	BMW	6 Series	650i	Convertible	automatic	wba9p9b57e4160262	CA	3	10736	black	black	the hertz corporation
28	27	2015	Hyundai	Sonata	SE	Sedan	automatic	5npe2af4f0832482	CA	3	9281	silver	gray	enterprise vehicle excha
29	28	2015	Volvo	XC70	T6	Wagon	automatic	va4902bn3f198103	CA	4	16508	brown	brown	volvo na rep/world omni
30	29	2015	Volvo	XC70	T6	Wagon	automatic	va4902bn3f1989531	CA	4	12725	beige	beige	volvo na rep/world omni
31	30	2015	Volvo	XC70	DriveES	SUV	automatic	va4902bn3f2403133	CA	4	11278	gray	black	audition san diego
32	31	2014	Chevrolet	Camaro	LT	Coupe	automatic	2t1fbf31e9238302	CA	4	11874	gray	black	midway hft fleet/ars
33	32	2014	Chevrolet	Cruze	1LT	Sedan	automatic	1g1pc5sb272109471	CA	4	2788	gray	black	enterprise vehicle excha
34	33	2015	Audi	A6	2.5	Sedan	automatic	wvauff3f3030345	CA	4	9557	white	black	audition vjeo
35	34	2015	Kia	Sorento	UX	SUV	automatic	5xvt6a69f5665226	CA	5	13757	red	black	kia motors america, inc
36	35	2014	Audi	S5	Premium Plus quattro	Coupe	automatic	wvaugcfvrea001689	CA	3	9875	black	black	fratelli investment group
37	36	2015	Volvo	S60	T5	Sedan	automatic	5xvt6a69f5660943	CA	4	12862	gray	black	kia motors america, inc
38	37	2014	Buick	Verano	Convenience Group	Sedan	automatic	1g4p83k4e4175320	CA	2	19531	gray	gray	enterprise vehicle excha
39	38	2015	Chevrolet	Suburban	LT	SUV	automatic	1n9skk3cfr187901	CA	4	1426	black	black	midway hft fleet/ars
40	39	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3a31c0a41163531	CA	4	111	black	black	financial services reman
41	40	2014	BMW	M5	Base	Sedan	automatic	wba9f5c1e59593069	CA	3	16360	black	black	the hertz corporation
42	41	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3a31c0a41163531	CA	4	9027	white	black	financial services reman

Moreover, we observed unknown values in columns “color” and “interior” which we changed to blanks.

Unknown Values Changed to Blanks

1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20		21		22		23		24		25		26		27		28		29		30		31		32		33		34		35		36		37		38		39		40		41		42		43		44		45		46		47		48		49		50		51		52		53		54		55		56		57		58		59		60		61		62		63		64		65		66		67		68		69		70		71		72		73		74		75		76		77		78		79		80		81		82		83		84		85		86		87		88		89		90		91		92		93		94		95		96		97		98		99		100		101		102		103		104		105		106		107		108		109		110		111		112		113		114		115		116		117		118		119		120		121		122		123		124		125		126		127		128		129		130		131		132		133		134		135		136		137		138		139		140		141		142		143		144		145		146		147		148		149		150		151		152		153		154		155		156		157		158		159		160		161		162		163		164		165		166		167		168		169		170		171		172		173		174		175		176		177		178		179		180		181		182		183		184		185		186		187		188		189		190		191		192		193		194		195		196		197		198		199		200		201		202		203		204		205		206		207		208		209		210		211		212		213		214		215		216		217		218		219		220		221		222		223		224		225		226		227		228		229		230		231		232		233		234		235		236		237		238		239		240		241		242		243		244		245		246		247		248		249		250		251		252		253		254		255		256		257		258		259		260		261		262		263		264		265		266		267		268		269		270		271		272		273		274		275		276		277		278		279		280		281		282		283		284		285		286		287		288		289		290		291		292		293		294		295		296		297		298		299		300		301		302		303		304		305		306		307		308		309		310		311		312		313		314		315		316		317		318		319		320		321		322		323		324		325		326		327		328		329		330		331		332		333		334		335		336		337		338		339		340		341		342		343		344		345		346		347		348		349		350		351		352		353		354		355		356		357		358		359		360		361		362		363		364		365		366		367		368		369		370		371		372		373		374		375		376		377		378		379		380		381		382		383		384		385		386		387		388		389		390		391		392		393		394		395		396		397		398		399		400		401		402		403		404		405		406		407		408		409		410		411		412		413		414		415		416		417		418		419		420		421		422		423		424		425		426		427		428		429		430		431		432		433		434		435		436		437		438		439		440		441		442		443		444		445		446		447		448		449		450		451		452		453		454		455		456		457		458		459		460		461		462		463		464		465		466		467		468		469		470		471		472		473		474		475		476		477		478		479		480		481		482		483		484		485		486		487		488		489		490		491		492		493		494		495		496		497		498		499		500		501		502		503		504		505		506		507		508		509		510		511		512		513		514		515		516		517		518		519		520		521		522		523		524		525		526		527		528		529		530		531		532		533		534		535		536		537		538		539		540		541		542		543		544		545		546		547		548		549		550		551		552		553		554		555		556		557		558		559		560		561		562		563		564		565		566		567		568		569		570		571		572		573		574		575		576		577		578		579		580		581		582		583		584		585		586		587		588		589		590		591		592		593		594		595		596		597		598		599		600		601		602		603		604		605		606		607		608		609		610		611		612		613		614		615		616		617		618		619		620		621		622		623		624		625		626		627		628		629		630		631		632		633		634		635		636		637		638		639		640		641		642		643		644		645		646		647		648		649		650		651		652		653		654		655		656		657		658		659		660		661		662		663		664		665		666		667		668		669		670		671		672		673		674		675		676		677		678		679		680		681		682		683		684		685		686		687		688		689		690		691		692		693		694		695		696		697		698		699		700		701		702		703		704		705		706		707		708		709		710		711		712		713		714		715		716		717		718		719		720		721		722		723		724		725		726		727		728		729		730		731		732		733		734		735		736		737		738		739		740		741		742		743		744		745		746		747		748		749		750		751		752		753		754		755		756		757		758		759		760		761		762		763		764		765		766		767		768		769		770		771		772		773		774		775		776		777		778		779		780		781		782		783		784		785		786		787		788		789		790		791		792		793		794		795		796		797		798		799		800		801		802		803		804		805		806		807		808		809		810		811		812		813		814		815		816		817		818		819		820		821		822		823		824		825		826		827		828		829		830		831		832		833		834		835		836		837		838		839		840		841		842		843		844		845		846		847		848		849		850		851		852		853		854		855		856		857		858		859		860		861		862		863		864		865		866		867		868		869		870		871		872		873		874		875		876		877		878		879		880		881		882		883		884		885		886		887		888		889		890		891		892		893		894		895		896		897		898		899		900		901		902		903		904		905		906		907		908		909		910		911		912		913		914		915		916		917		918		919		920		921		922		923		924		925		926		927		928		929		930		931		932		933		934		935		936		937		938		939		940		941		942		943		944		945		946		947		948		949		950		951		952		953		954		955		956		957		958		959		960		961		962		963		964		965		966		967		968		969		970		971		972		973		974		975		976		977		978		979		980		981		982		983		984		985		986		987		988		989		990		991		992		993		994		995		996		997		998		999		1000		1001		1002		1003		1004		1005		1006		1007		1008		1009		1010		1011		1012		1013		1014		1015		1016		1017		1018		1019		1020		1021		1022		1023		1024		1025		1026		1027		1028		1029		1030		1031		1032		1033		1034		1035		1036		1037		1038		1039		1040		1041		1042		1043		1044		1045		1046		1047		1048		1049		1050		1051		1052		1053		1054		1055		1056		1057		1058		1059		1060		1061		1062		1063		1064		1065		1066		1067		1068		1069		1070		1071		1072		1073		1074		1075		1076		1077		1078		1079		1080		1081		1082		1083		1084		1085		1086		1087		1088		1089		1090		1091		1092		1093		1094		1095		1096		1097		1098		1099		1100		1101		1102		1103		1104		1105		1106		1107		1108		1109		1110		1111		1112		1113		1114		1115		1116		1117		1118		1119		1120		1121		1122		1123		1124	
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	-----	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--	------	--

Splitting of column "saledate"

Page | 7

Finally, we changed blanks with N/A values for the variables we are planning to use as Dimensions on our Fact Table later.

Replacement of Blanks by N/A Values

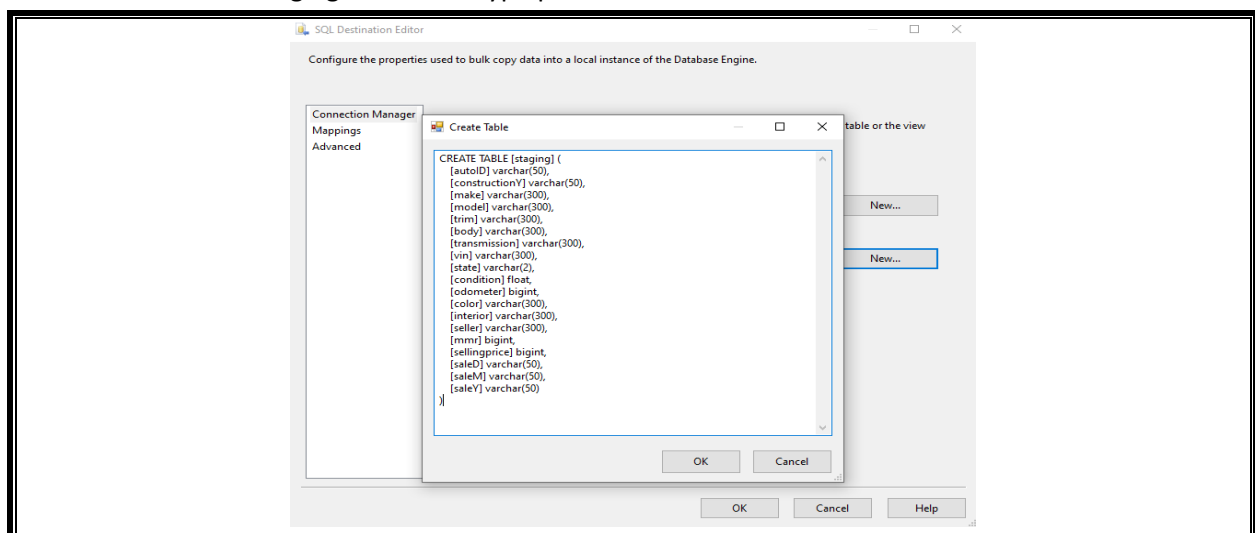
[illegible]

Data Import

First, we created a relational database in SQL Server called “used_cars”. Then, in Visual Studio, we created a Data Flow Task called “Import CSV” which includes a Flat-file Source and an SQL Server Destination. The Flat-file Source was imported with the cleaned CSV file of our dataset.

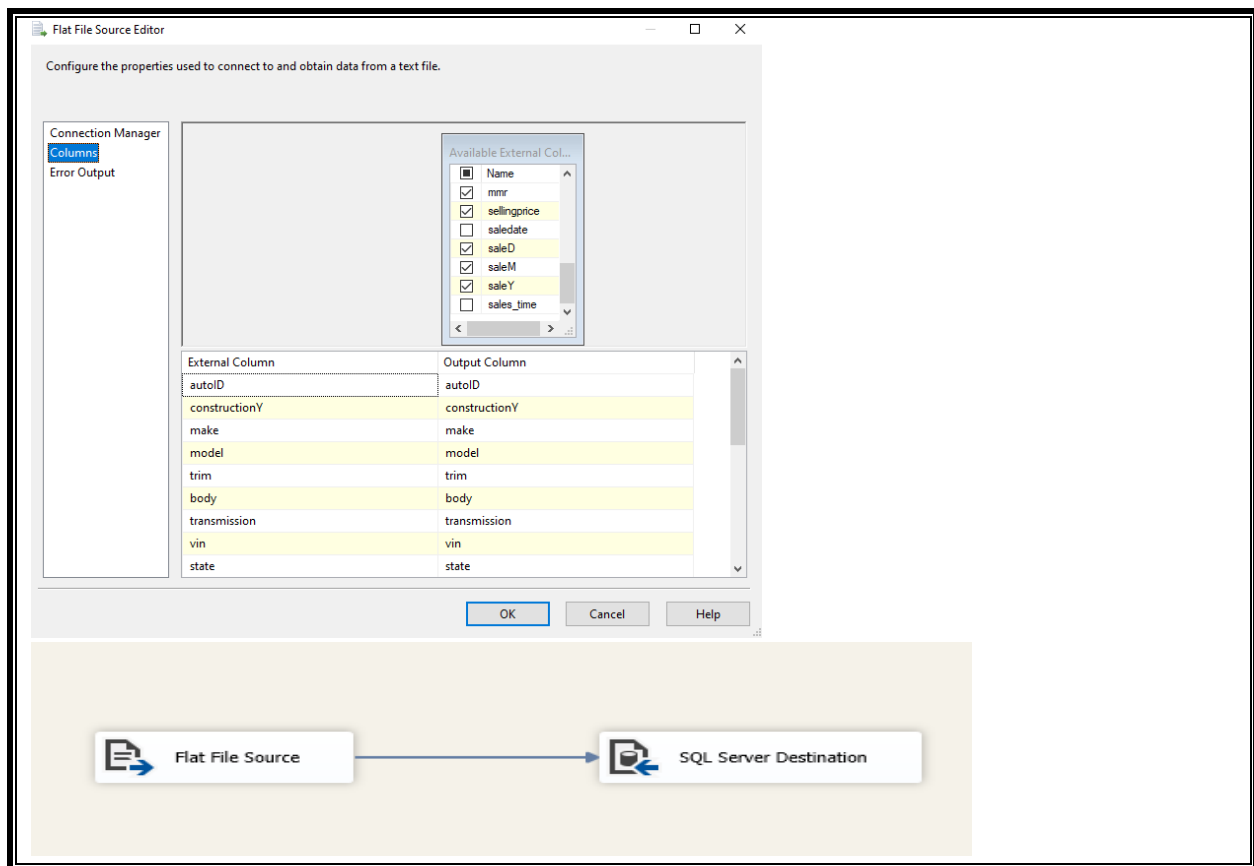
We connected SQL Server Destination to the Flat-file Source in order to connect our CSV directly to the database “used_cars” in SQL Server and used it to create a new table called “staging” in our database. By creating table “staging”, we declared the data types of our columns.

Creation of Table “staging” and Data Type per Column Declaration



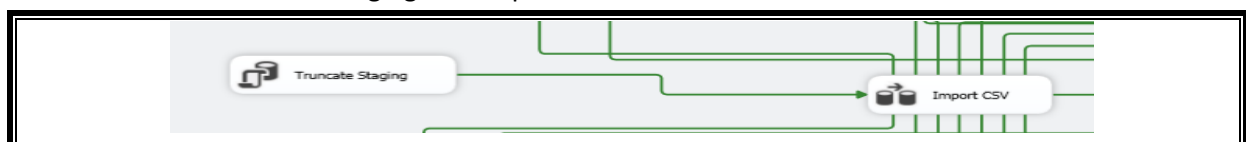
Finally, we chose the columns we will use in our report, connected the Flat-file Source to the SQL Server Destination and imported the CSV file to the “staging” table.

Selection of Columns and Connection of Flat-file Source to SQL Server Destination

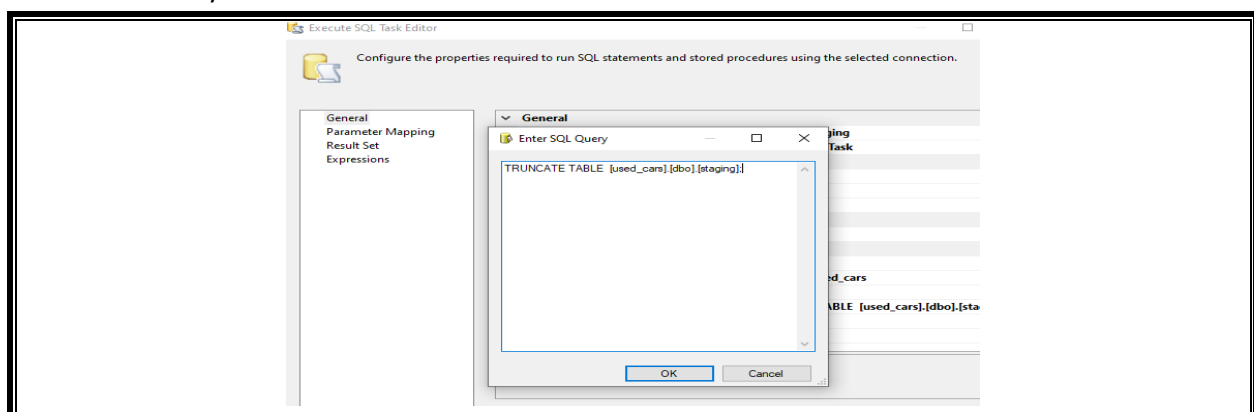


Furthermore, we connected the “Import CSV” process, with an SQL Execute Task, that would truncate our “staging” table each time we execute the data flow.

Connection of “Truncate Staging” to “Import CSV”



Truncate Query Execution



Afterwards, we defined the dimensions and measures for the fact table. After consideration, we decided that the following columns will be used as dimensions:

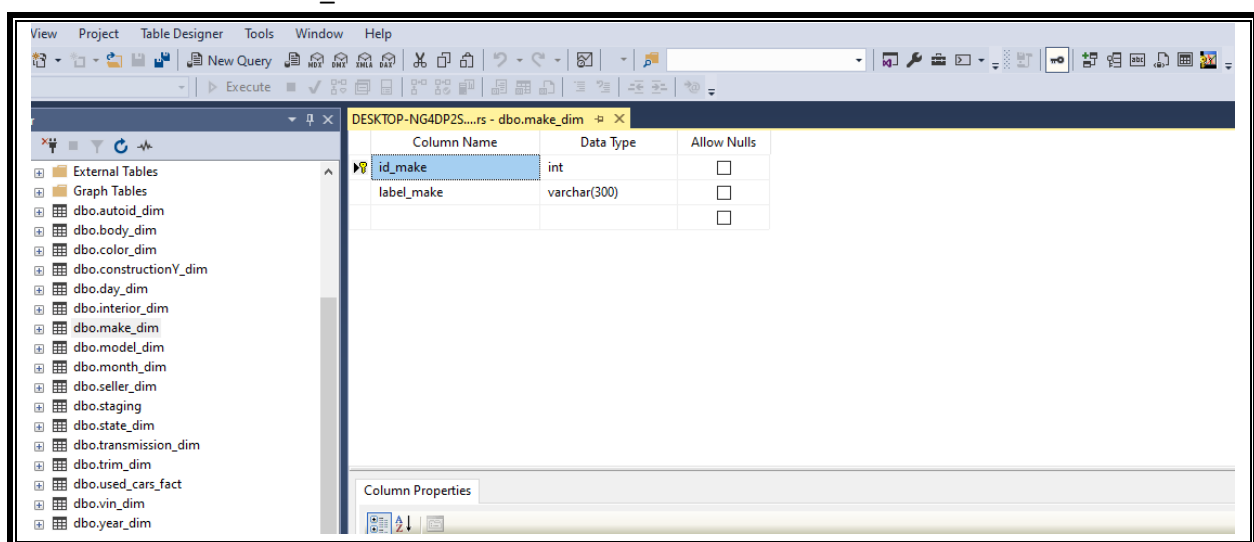
- autoid
- constructionY
- make
- model
- trim
- body
- transmission
- color
- interior
- state
- vin
- seller
- saleD
- saleM
- saleY

and the following will be used as measures:

- sellingprice
- mmr
- condition
- odometer

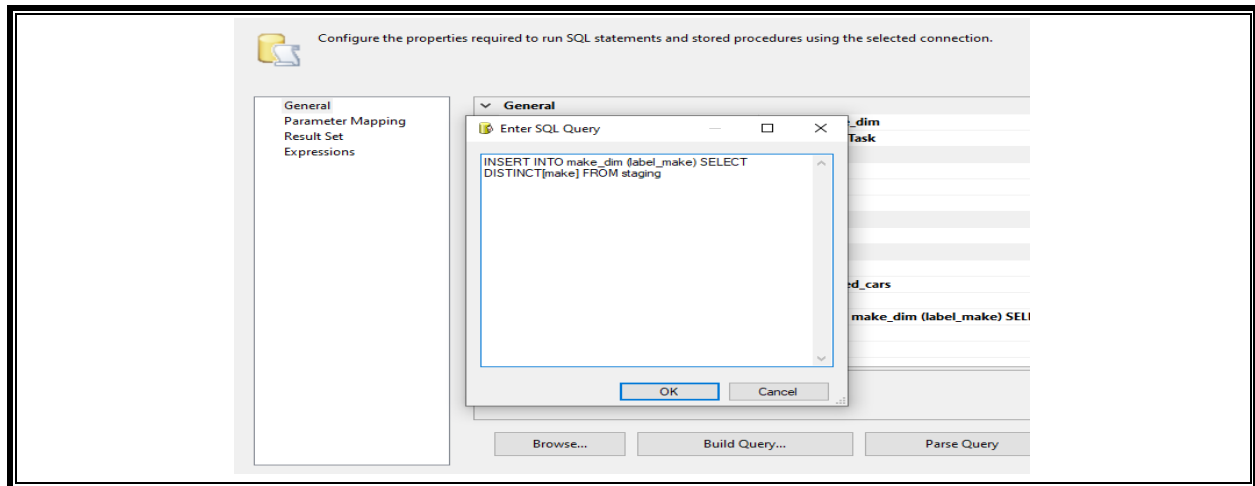
To implement that, we created a new table for each dimension. In the following image we demonstrate the construction of a dimension table with a primary key and a column called “label_make” to insert the values of column make. The same procedure was followed for all the other Dimensions.

Creation of Table “make_dim”



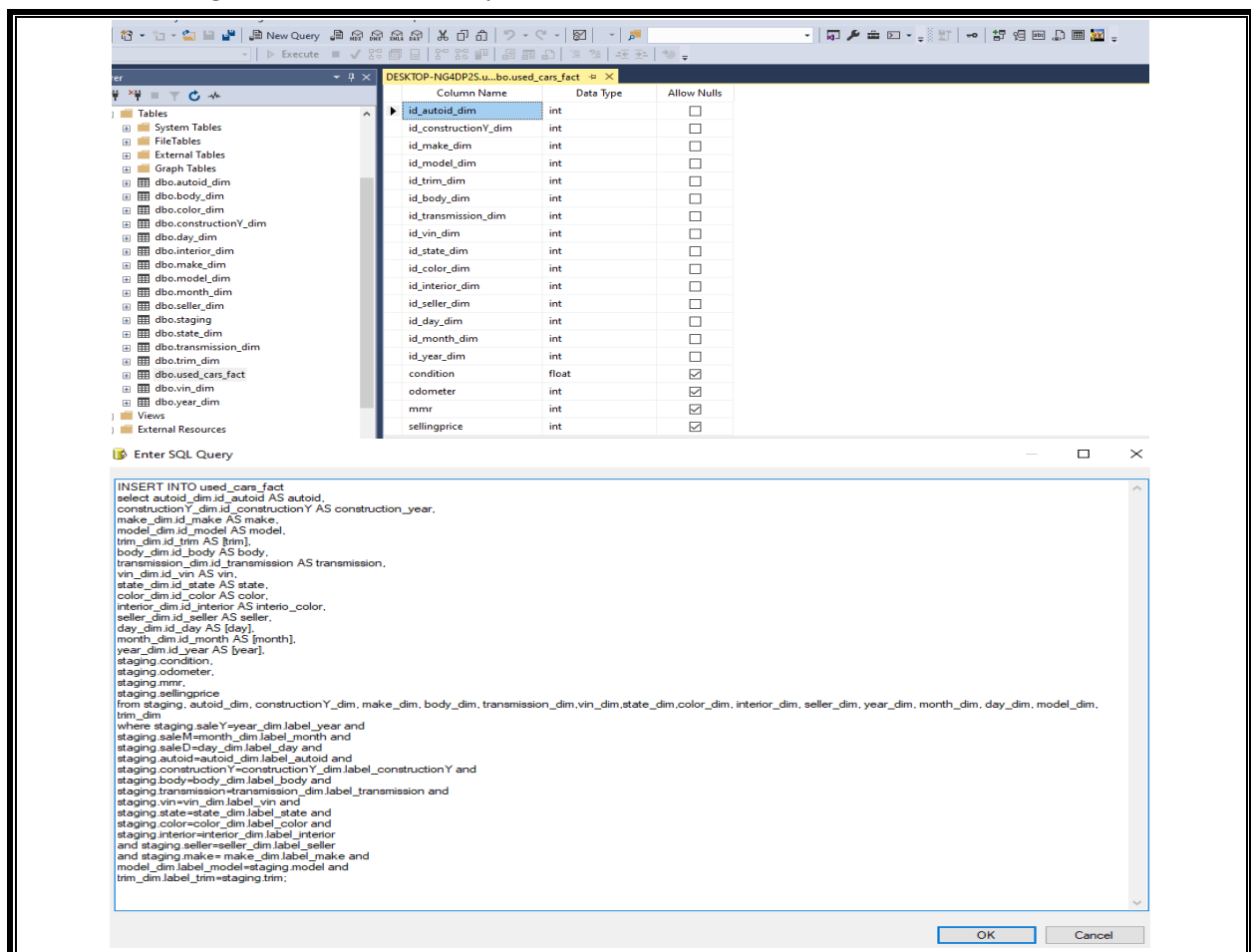
Afterwards, in Visual Studio, we created an Execute SQL Task for each dimension which contains the SQL Query that inserts values into our dimension tables and connected each one of them with the “Import CSV”.

SQL Insert Query for Table “make_dim”



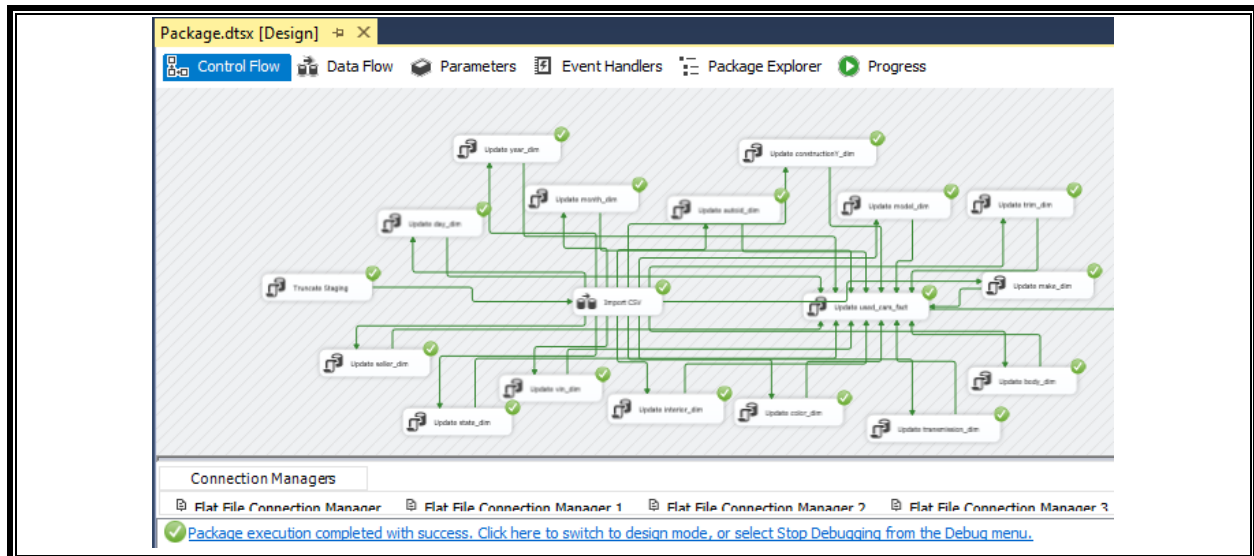
Finally, we created the fact table called “used_cars_fact” which contains as foreign keys the primary keys of each dimension and the 4 measures we described before.

Fact Table Design and SQL Insert Query



Afterwards, in Visual Studio, we created an Execute SQL Task for fact table which contains the SQL Query that inserts values into our fact table and connects it with each one dimension's SQL Execute Task. The final running structure of our control flow in Visual Studio is the following:

Connection of "Update used_cars_fact" to the other Execute SQL Tasks



After executing the control flow, we can observe that every table is filled with values. By running a select statement in MSSQL we can assure that. We demonstrate the results of one dimension and of the fact table.

Query 1: SELECT * FROM [used_cars].[dbo].[used_cars_fact]

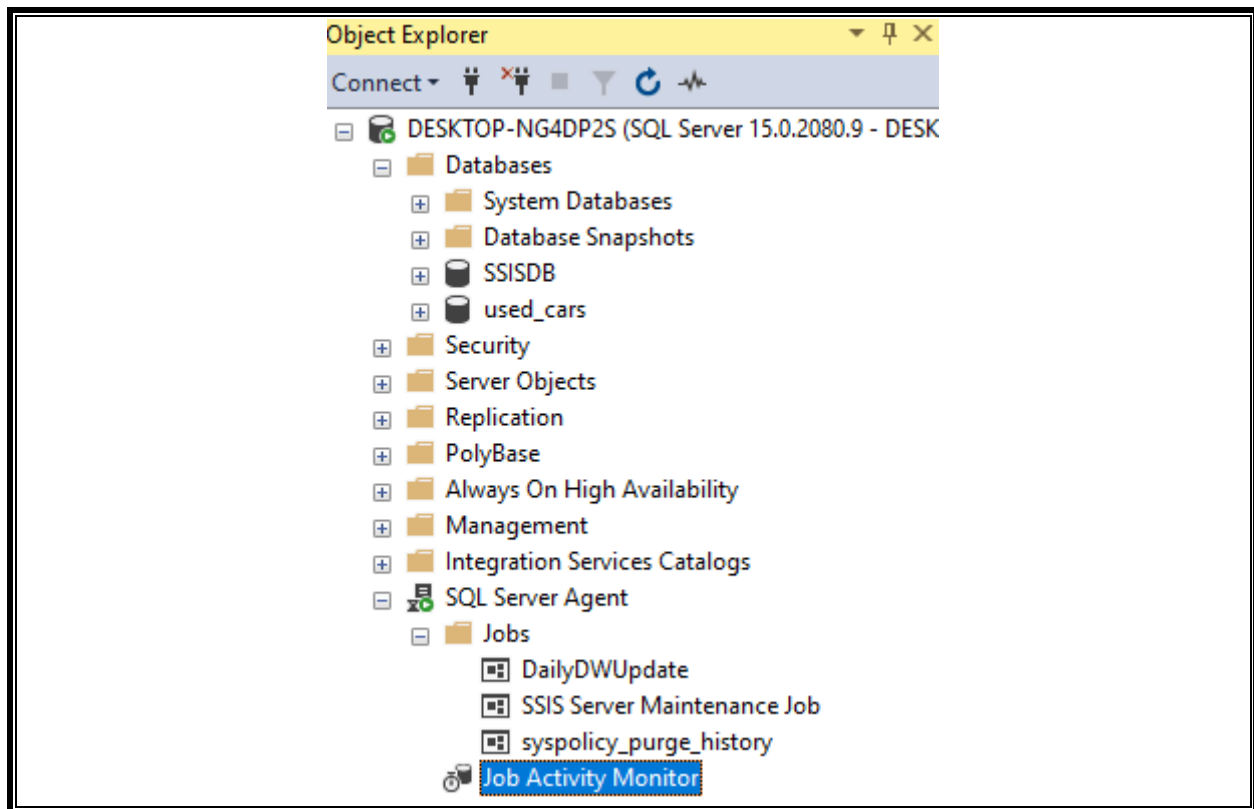
id_autoid_dim	id_construction_dim	id_make_dim	id_model_dim	id_trim_dim	id_body_dim	id_transmission_dim	id_vin_dim	id_state_dim	id_color_dim	id_interior_dim	id_seller_dim	id_day_dim	id_month_dim	id_year_dim	condition	odometer	mnr	sellingprice
31	316466	18	27	82	240	37	2	200759	11	1	13839	22	2	2	34	124696	3775	3500
32	316467	18	27	82	120	37	2	181666	10	1	1179	26	2	2	19	118825	3325	2750
33	316468	18	28	491	1517	37	2	417593	10	8	245	26	2	2	3	123638	3375	5600
34	316470	18	27	86	1410	33	2	372953	34	1	11676	26	2	2	21	122115	8150	8300
35	316471	18	3	40	528	37	2	545815	18	1	13992	1	2	2	2	1	5275	15750
36	316472	18	27	557	1113	33	3	375339	13	1	11676	18	2	2	41	88972	7100	9800
37	316474	18	27	526	1410	33	2	444183	33	7	13992	26	2	2	19	135372	5350	3000
38	316475	18	27	526	1473	33	2	444655	28	1	4936	7	2	2	37	113695	5750	7300

Query 2: SELECT * FROM [used_cars].[dbo].[make_dim]

id_make	label_make
1	Acura
2	Audi
3	Aston Martin
4	Audi
5	Bentley
6	BMW
7	Buick
8	Cadillac
9	Chevrolet
10	Chrysler
11	Daimler
12	Dodge
13	Ferrari
14	FIAT
15	Ford
16	Ford
17	Gao
18	GMAC
19	Honda
20	HUMMER
21	Honda
22	Infiniti

Finally, we deployed an SSIS Package which is responsible to execute the Visual from the SQL Server daily at 2 am.

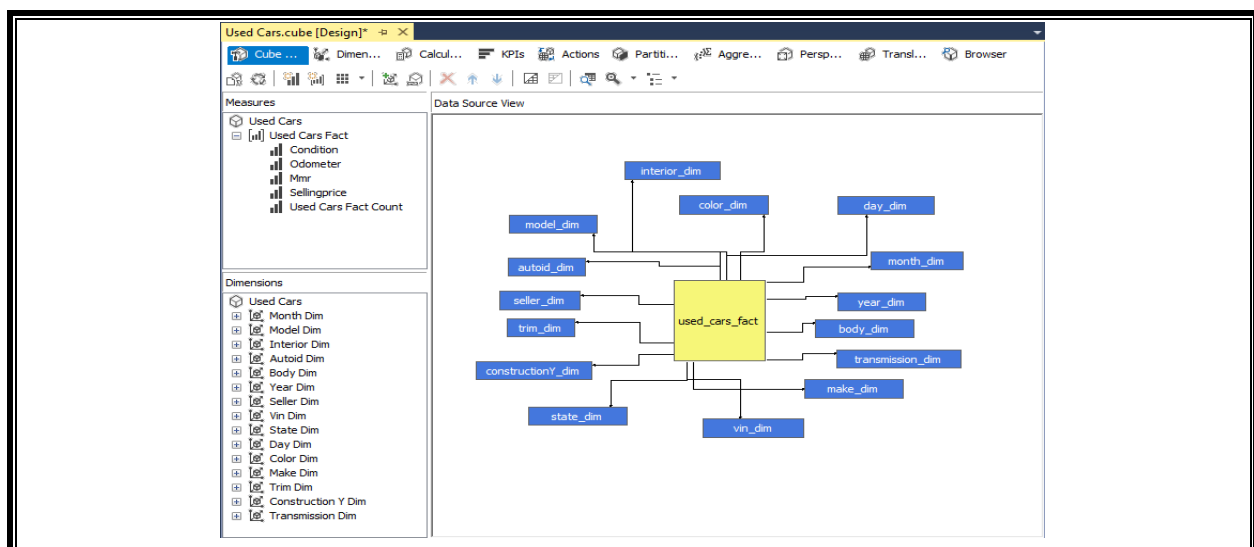
Job Activity Monitor Creation



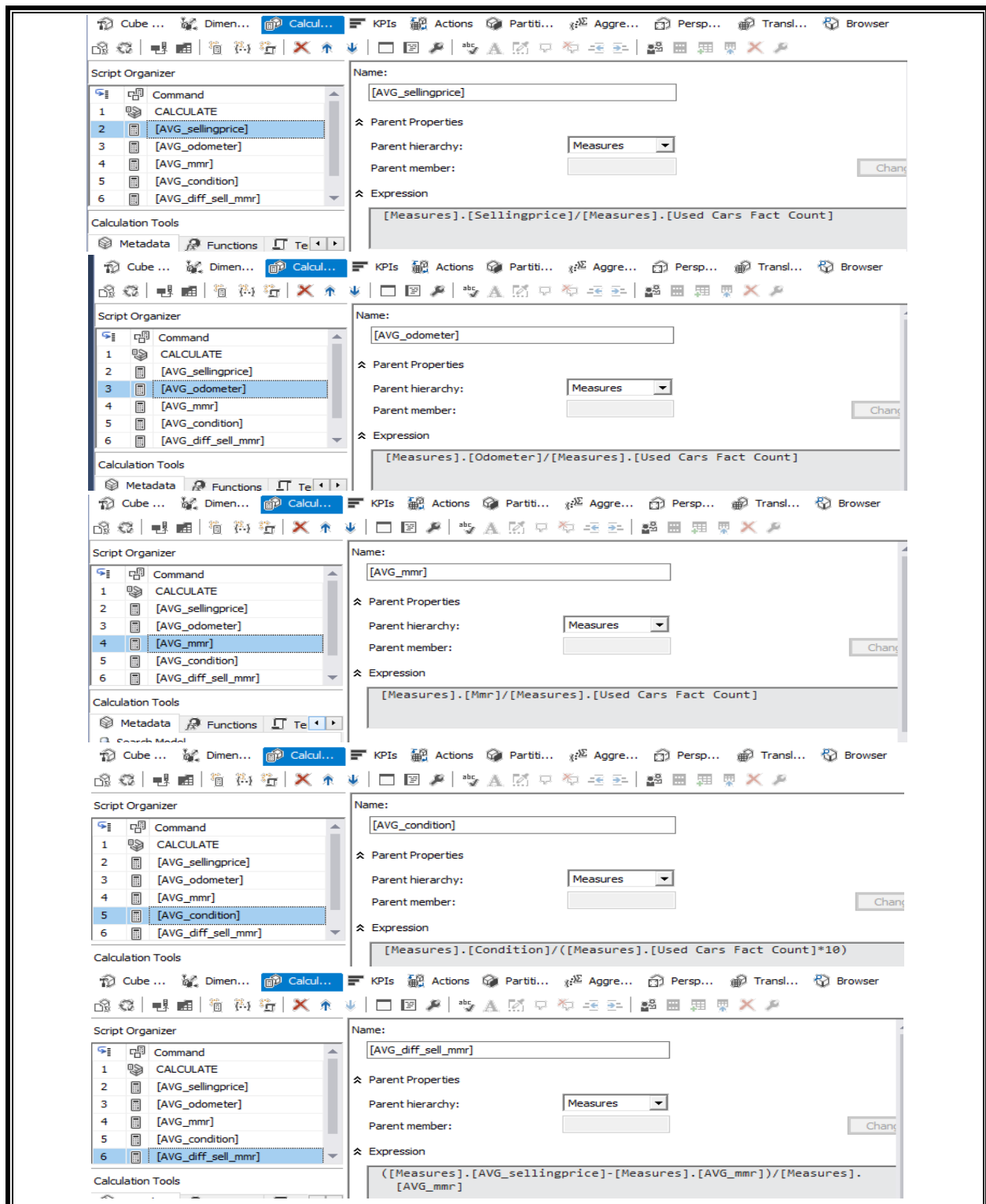
Data Cube Deployment

We created a Multidimensional Analysis Services project in SSIS and connected it to our SQL Server relational database to create a cube with the help of Cube Wizard. Our final cube Schema is the following:

Cube Visualization

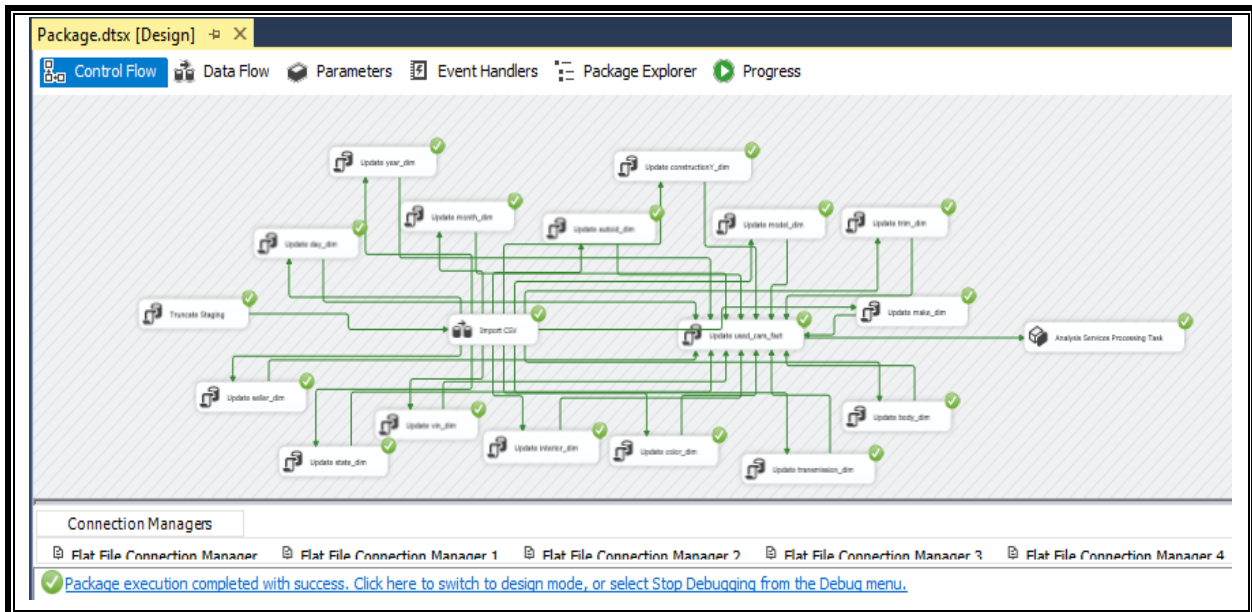


We matched dimensions and measures of our database to the Cube and calculated some extra measures (the averages of our already existing measures) to use them later in our visualizations.



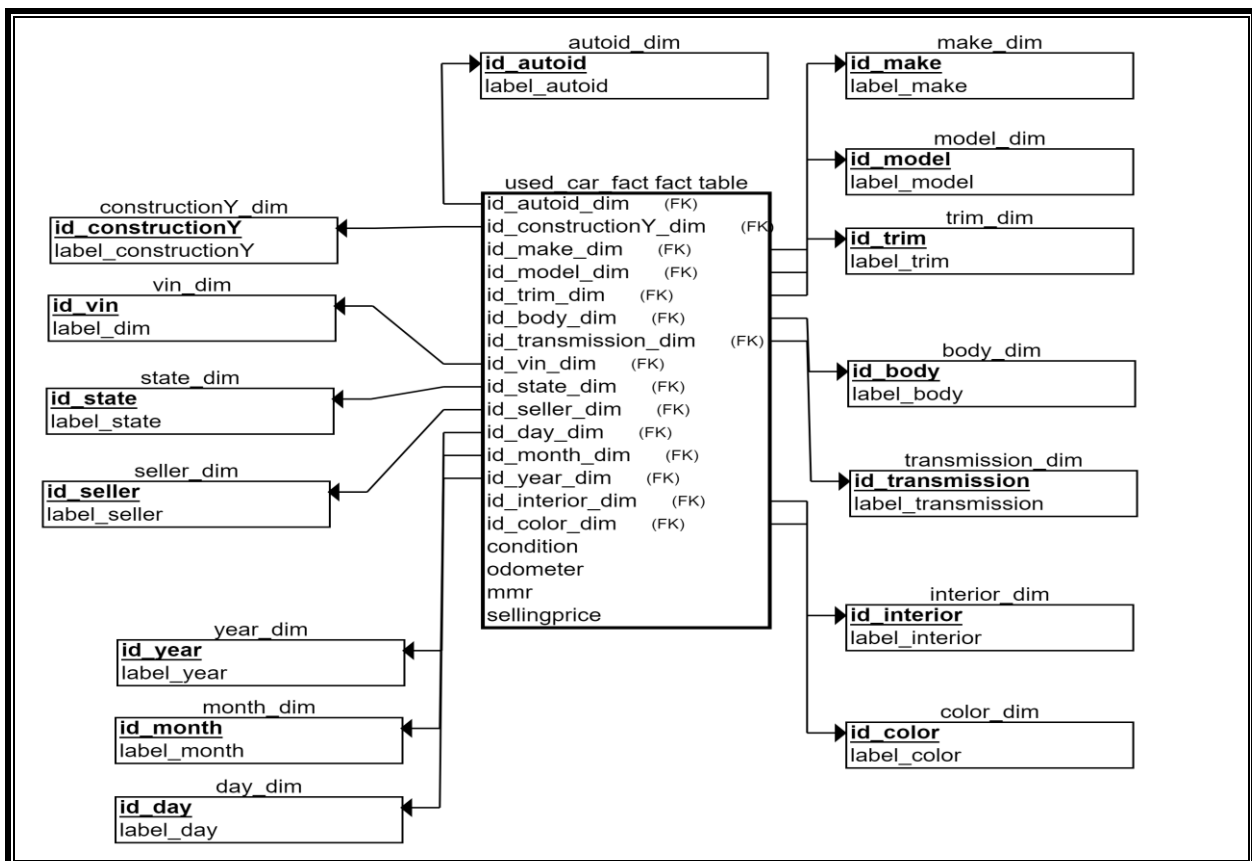
We connected the Cube to our Visual flow via an Analysis Services Processing Task.

Connection of ASPT to “Update used_cars_fact”



Finally, we connected our Cube to Power BI to visualize our data. The star schema of our database is the following:

Star Schema

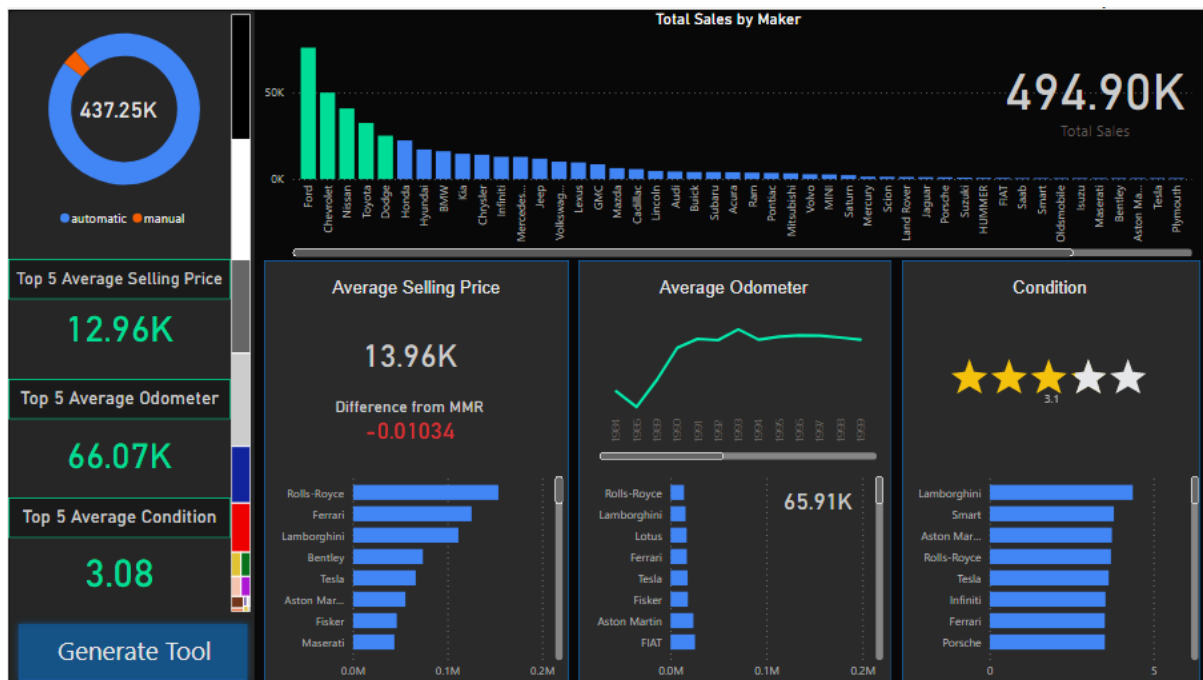


Business Analysis

Step 1

Using Power BI, we created a visualization about the car auctions market and presented the most important features of it for the first 7 months of 2015. On the top, we created a stacked column chart which presented the Total and Top 5 Sales by Maker. The Total Sales reached 494.90k and the Top 5 Makers by Sales are Ford (76k), Chevrolet(50k), Nissan(40k), Toyota(32k) and Dodge(25k). The Top 5 constitutes 45% of the market's Total Sales and are mostly USA and Japan produced cars which indicates that US citizens prefer the domestic car makers market but also, the well-established names of Japanese cars' brands. Also, this indicates that it is easier for Auto+ to be supplied cars created by the Top 5 makers as there is greater supply and competitive prices for them in the US used cars market. Moreover, it is observed that, from the Total of Sales, the vast majority consists of automatic cars (437.25k) against manual cars (15.07k) which indicates the preference of US citizens to automatic cars. Finally, the most preferred colors in descending order are black, white, silver, gray, blue and red.

Market Analysis Report

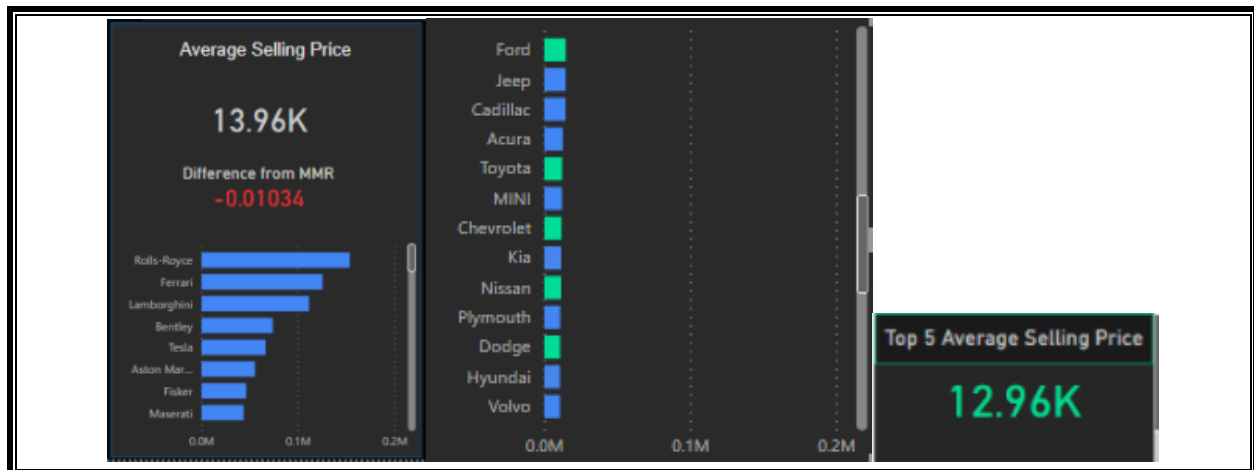


Furthermore, we studied the sales market by 3 competitive factors:

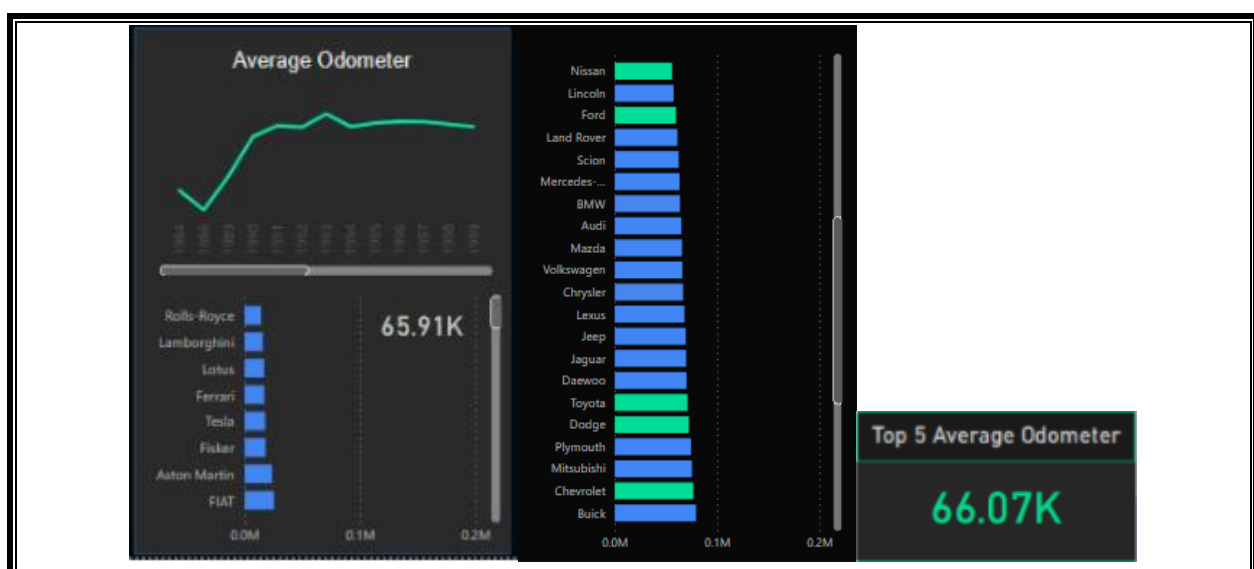
- Average Selling Price
- Average Odometer
- Average Condition

Average Selling Price: In this tab, is presented the Average Selling Price of the market and its difference from the MMR indicator while on the left is presented the Average Selling Price for the Top 5 makers. It is observed that the market values luxury cars the most as Rolls-Royce, Ferrari and Lamborghini are the highest priced cars while the Top 5 Sales cars are found mostly in the mid of the list. The Average Selling Price of the Top 5 is \$12.96k and is lower from the general

Average Price by \$1k. This indicates that the Top 5 are in favor of the average US citizen in terms of average price and that selling price plays a critical role in the customer's final decision. Finally, the difference between the Average Selling Price and the MMR indicator for the Total of the market, as well as, per maker presents almost zero difference which indicates low goodwill from the auction of used cars.



Average Odometer: In this Tab, is presented the Average Odometer the cars of used cars market have completed and their correlation to the year the cars were constructed while on the left is presented the Average Odometer for the Top 5 makers. It is observed that the Average Odometer is highly negatively correlated to the construction year as it is getting higher while the construction year is ascending. There is a small number of outliers from 1984 and 1986 with very low mileage in contrast to their construction year but we can assume they are abandoned or unused cars. The Top 5 presents 66.07k on Average which is higher from the general average by 0.16k. The customers, on average, prefer cars close to 65.91k miles (the Average Odometer value) which indicates that they prefer cars with medium mileage. Chevrolet, Dodge and Toyota are most preferred as the construction year ascends, and this is why the Average Odometer of Top 5 is higher than the Average Odometer of general.



Average Condition: In this Tab, is presented the Average Condition of the cars while on the left is presented the Average Condition for the Top 5 makers. It is observed that the average car is in condition 3.1 in a scale 1-5. The Top 5 cars present an average of 3.08 Condition which is lower to the Average Condition by 0.02. This indicates that Customers prefer cars with at least a good rating of Condition.



In conclusion, Customers, on average, prefer cars that balance between mileage and condition but it is observed that a higher priority is given to selling price. The Customers are willing to purchase a car with average odometer and condition close to market's average, while they prefer cars lower to average selling price by at least 1k.

Step 2

To get a more detailed view on the top 5 Makers' cars, we created a tool that allows the user to compare models and trims from the Top 5 Makers. Users can choose their makers on the top of the page, models and trims on the lists left and right of the page and color on the bottom of the page. Additionally, they can observe the distribution of prices per maker, model and trim from the bar charts on the left and right bottom of the page. In this way, users can easily identify key differences on the selling price, mileage and condition and get supplied makers, models and trims that are closer to their customer preferences. Additionally, the user is able to visualize their car's color choosing from a list of top 6 more applied colors. We used this tool in order to choose the best models from each maker and consult our customer on the best fit models they need to be supplied for their entry into the market.



Similarly to Step 1, we observed the models of higher selling amount per maker and compared them to the average selling price, average odometer and average condition of the market. It is observed that:

- For maker **Chevrolet** the top 2 models based on Sales are Impala (average selling price = 8.23k, average odometer = 80.41k, average condition = 2.5/5) and Malibu (average selling price = 9.22k, average odometer = 67.61k, average condition = 2.89/5).
- For maker **Dodge** the top 2 models based on Sales are Grand Caravan (average selling price = 11.62k, average odometer = 64.70k, average condition = 2.59/5) and Charger (average selling price = 13.29k, average odometer = 63.24k, average condition = 2.60/5).
- For maker **Ford** the top 2 models based on Sales are F150 (average selling price = 19.20k, average odometer = 70.83k, average condition = 3.25/5) and Fusion (average selling price = 12.46k, average odometer = 46.87k, average condition = 3.51/5).
- For maker **Nissan** the top 2 models based on Sales are Altima (average selling price = 11.51k, average odometer = 49.04k, average condition = 3.25/5) and Maxima (average selling price = 13.31k, average odometer = 62.79k, average condition = 3.09/5).
- For maker **Toyota** the top 2 models based on Sales are Camry (average selling price = 11.24k, average odometer = 66.05k, average condition = 3.08/5) and Corolla (average selling price = 9.34k, average odometer = 64.07k, average condition = 2.91/5).

We created a scorecard to evaluate the Top Models:

	Average Selling Price	Average Odometer	Average Condition	✓	X	Total Total(✓)-Total(X)
Chevrolet Impala	✓ ✓	X	X	2	2	0
Chevrolet Malibu	✓ ✓	✓	X	3	1	2

Dodge Grand Caravan	✓	✓	X	2	1	1
Dodge Charger	X	✓	X	1	2	-1
Ford F150	XX	X	✓	1	3	-2
Ford Fusion	✓	✓	✓	3	0	3
Nissan Altima	✓	✓	✓	3	0	3
Nissan Maxima	X	✓	✓	2	1	1
Toyota Camry	✓	✓	✓	3	0	3
Toyota Corolla	✓ ✓	✓	X	3	1	2

✓ : Model is close to average Customer's preference

X: Model is far from average Customer's preference

✓ ✓ : Model is close to average Customer's preference and preference plays a critical role on Customer's final decision*

XX: Model is far from average Customer's preference and preference plays a critical role on Customer's final decision*

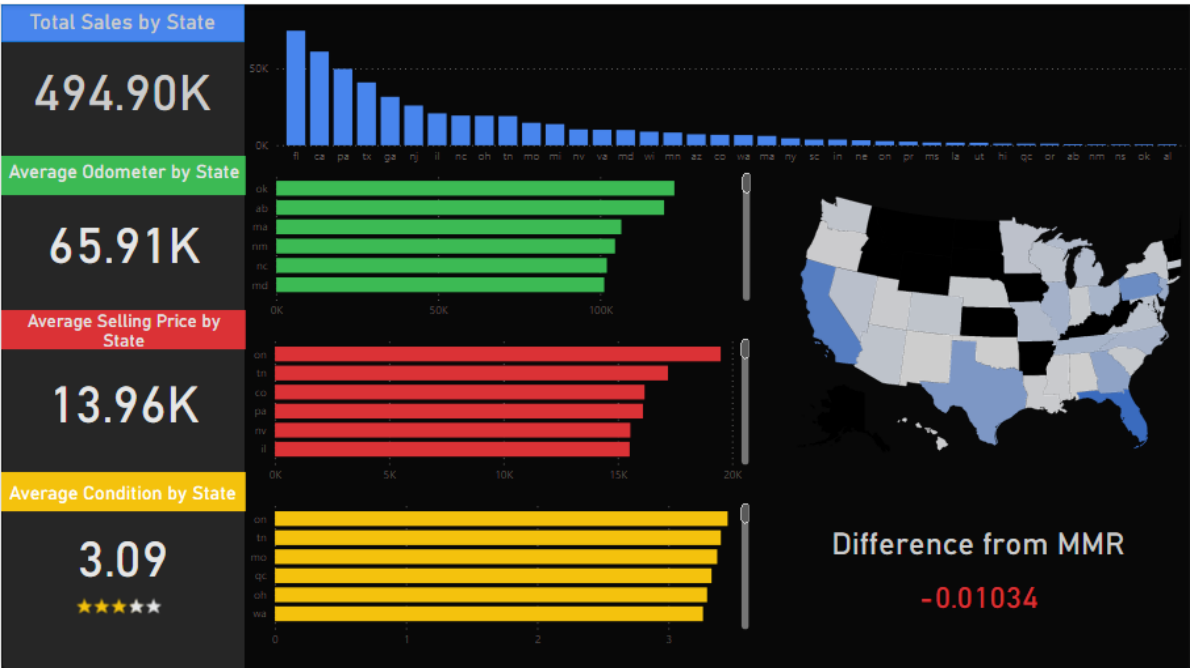
*Low selling price plays a critical role on Customer's final decision as we concluded in Step 1

In conclusion, it is suggested that Auto+ could get supplied of models Chevrolet Malibu, **Dodge** Grand Caravan, **Ford** Fusion, **Nissan** Altima, **Nissan** Maxima, **Toyota** Camry and **Toyota** Corolla as they perform the highest on our scoreboard and it is believed that **Ford** Fusion, **Nissan** Altima and **Toyota** Camry will outperform the other models on level of Sales as the higher and most balanced performers in terms of average selling price, average odometer and average condition of our scorecard.

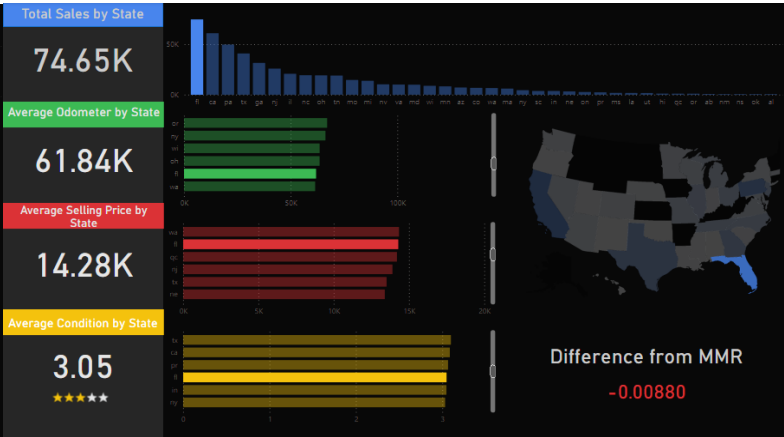
Step 3

To decide about the best location to open the store, we created a report based on the Total Sales per State. It is observed that California State is the best choice for our client. While Florida is observed to be the highest on Total Sales among other states (74.65k), California State presents a high number of Total Sales (61.03k) and lower performance on our key factors in comparison to the car models we advised our customer to supply. More in detail, California's average selling price is 14.84k in comparison to our stock's selling price of 11.24k, California's average odometer is 67.24 in comparison to our stock's odometer of 60.16k and California's average condition is 3.09 in comparison to our stock's condition of 3.04. Moreover, the average MMR of California is +0.00261 which indicates a small profit on each car sale based on California's average selling

price. The low average selling price of our client’s stock is giving them the opportunity to sell on higher price and on a margin of +3.6k until California's market average selling price is reached.



CALIFORNIA



FLORIDA