

Used Car Sales Analysis

XINAN WANG [002916472]

KEWEI SHI [001359675]

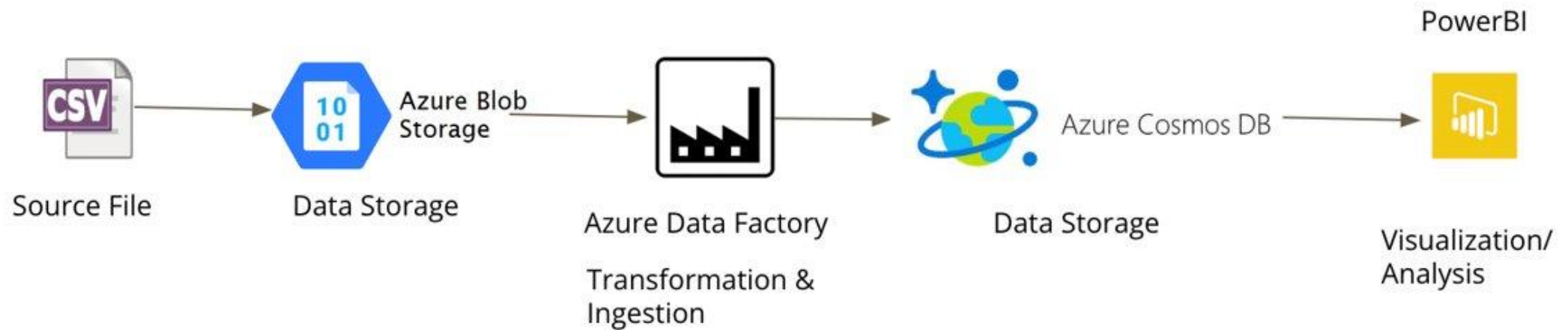
ZHIQING WANG [002916484]

JIAYUE HAN [002965004]

Objectives

- STUDY USED CAR SALES DATA IN US
- ANALYZE USED CAR DATA AND DRIVE INSIGHTS ON USED CAR SALES TREND
- DISCOVER TRAITS OF POPULAR USED CARS MODELS, BRANDS, ETC.,
- CREATE VISUALIZATION USING MICROSOFT POWERBI

Data Architecture Diagram



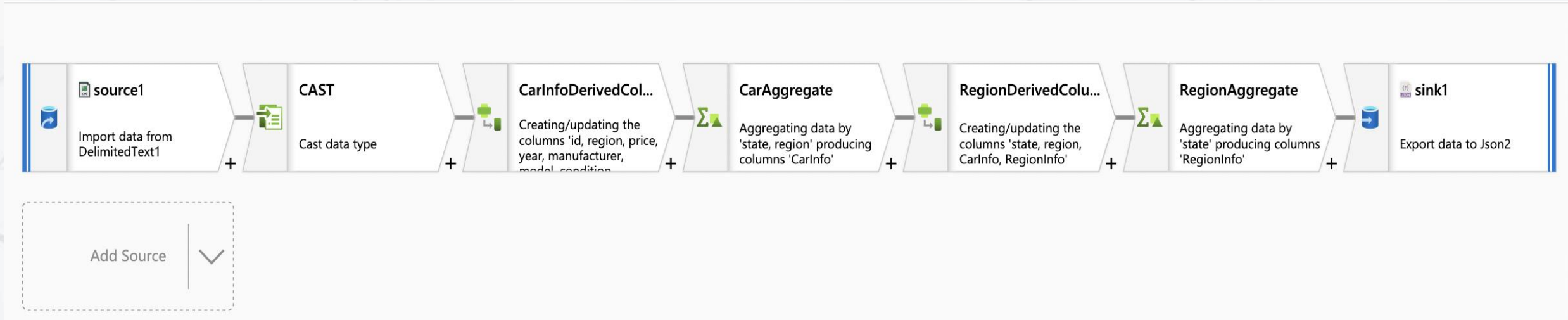
Source Data

id	region	price	year	manufacturer	model	condition	cylinders	fuel	odometer	title_status	transmission	VIN	drive	size	type	paint_color	state
7301607385	scranton / wilkes-barre	15990	2014.0	kia	sorento lx sport utility 4d	good		gas	93303.0	clean	other	5XYKT3A64EG443919	fwd		SUV	black	pa
7301607386	knoxville	11590	2014.0	chevrolet	spark ev 2lt hatchback	good		electric	61852.0	clean	other	KL8CL6S02EC412682	fwd		hatchback		tn
7301607388	beaumont / port arthur	11590	2014.0	chevrolet	spark ev 2lt hatchback	good		electric	61852.0	clean	other	KL8CL6S02EC412682	fwd		hatchback		tx
7301607390	green bay	30990	2017.0	lincoln	mkx reserve sport	good	6 cylinders	gas	23545.0	clean	automatic	2LMPJ8LR2HBL35260			SUV	blue	wi
7301607395	toledo	29990	2017.0	lincoln	mkx reserve sport	good	6 cylinders	gas	23545.0	clean	automatic	2LMPJ8LR2HBL35260			SUV	blue	oh
7301607396	flint	30990	2017.0	lincoln	mkx reserve sport	good	6 cylinders	gas	23545.0	clean	automatic	2LMPJ8LR2HBL35260			SUV	blue	mi
7301607402	cookeville	38990	2020.0	audi	q5 45 tfsi premium plus	good		gas	19067.0	clean	other	WA1BNAFYXL2013483			SUV	white	tn
7301607482	rochester	5500	2011.0		BLUE BIRD BLUE BIRD				95023.0		automatic	1BAKFC5A4CF287008					ny
7301607519	albany	200	2005.0		NONE	good		other	111000.0	clean	other					white	ny
7301607572	sierra vista	29990	2017.0	bmw	3-series			other	64677.0	clean	automatic		rwd		sedan		az
7301608024	boone	5500	2006.0		2006 SILVERADO 2500 HD	excellent	8 cylinders	gas	330500.0	clean	automatic	1GCHK29G96E238160	4wd	full-size	pickup	white	nc
7301608030	iowa city	700	1993.0	ford	ranger	fair	6 cylinders	gas	160000.0	clean	automatic		4wd			green	ia
7301608138	greenville / upstate	0	2017.0	chevrolet	colorado	good	4 cylinders	gas	27944.0	clean	automatic	1GCHSBEA2H1148845	rwd	full-size	truck	white	sc
7301608152	greenville / upstate	0	2019.0	dodge	grand caravan	good		gas	96887.0	clean	other	2C4RDGBG3KR614340		full-size	van	blue	sc
7301608251	richmond	4000	1994.0	mitsubishi	3000gt	good	6 cylinders	gas	149500.0	clean	manual		fwd		hatchback	custom	va
7301609037	chattanooga	12990	2011.0	acura	tsx sedan 4d	good		other	85757.0	clean	other	JH4CU2F63BC006531			sedan	black	tn
7301609038	altoona-johnstown	24590	2018.0	lincoln	mkz select sedan 4d	good		gas	16081.0	clean	automatic	3LN6L5C91JR615484	fwd		sedan	white	pa
7301609048	port huron	43990	2020.0	buick	enclave avenir sport	good	6 cylinders	gas	5217.0	clean	other	5GAERDKW5LJ239810	fwd		SUV		mi
7301609049	athens	38990	2017.0	acura	mdx sport hybrid sh-awd	good		other	9326.0	clean	other	5FRYD7H59HB001823			SUV	white	oh
7301609055	macon / warner robins	34590	2020.0	infiniti	qx50 luxe sport utility	good		other	3542.0	clean	other	3PCAJ5M11LF110384	fwd		hatchback	brown	ga
7301609056	albany	34590	2020.0	infiniti	qx50 luxe sport utility	good		other	3542.0	clean	other	3PCAJ5M11LF110384	fwd		hatchback	brown	ga
7301609062	charleston	24590	2018.0	lincoln	mkz select sedan 4d	good		gas	16081.0	clean	automatic	3LN6L5C91JR615484	fwd		sedan	white	wv
7301609064	kenosha-racine	38990	2020.0	audi	q5 45 tfsi premium plus	good		gas	19067.0	clean	other	WA1BNAFYXL2013483			SUV	white	wi
7301609071	columbus	59990	2019.0	audi	q8 premium sport utility 4d	good	6 cylinders	gas	8930.0	clean	automatic	WA1AFAF15KD015966			other	black	ga
7301609079	eastern CT	30590	2020.0	volvo	s60 t5 momentum sedan 4d	good		gas	12029.0	clean	other	7JR102FKXLG042696	fwd		sedan	red	ct
7301609081	winston-salem	49990	2020.0	cadillac	xt5 sport suv 4d	good	6 cylinders	gas	4322.0	clean	other	1GYKNGRS0LZ193218			other	white	nc
7301609082	evansville	29990	2017.0	lincoln	mkx reserve sport	good	6 cylinders	gas	23545.0	clean	automatic	2LMPJ8LR2HBL35260			SUV	blue	in
7301609083	eastern CT	40990	2019.0	acura	mdx advance pkg sport	good	6 cylinders	gas	12046.0	clean	other	5J8YD3H87KL002026	fwd		other	red	ct
7301609087	manhattan	14990	2015.0	honda	civic lx coupe 2d	good		gas	77726.0	clean	other	2HGFG3A55FH510450	fwd		coupe	red	ks
7301609094	utica-rome-oneida	34990	2020.0	cadillac	xt4 sport suv 4d	good		diesel	4174.0	clean	other	1GYFZFR46LF088296			hatchback	white	ny
7301609098	finger lakes	32990	2015.0	chevrolet	ss sedan 4d	good	8 cylinders	gas	84498.0	clean	automatic	6G3F15RW8FL107349	rwd		sedan	green	ny
7301609101	lima / findlay	41590	2020.0	audi	a6 45 tfsi premium plus	good		gas	27160.0	clean	other	WAUE8AF25LN004532			sedan	black	oh
7301609113	eastern CT	43990	2020.0	buick	enclave avenir sport	good	6 cylinders	gas	5217.0	clean	other	5GAERDKW5LJ239810	fwd		SUV		ct
7301609115	charlottesville	33590	2020.0		Genesis G70 2.0T Sedan 4D	good		other	786.0	clean	other	KMTG64LA9LU056364			sedan	white	va

Data in Blob Storage "Input"

```
1 id,region,price,year,manufacturer,model,condition,cylinders,fuel,odometer,title_status,transmission,VIN,
2 7301607385,scranton / wilkes-barre,15990,2014.0,kia,sorento lx sport utility 4d,good,,gas,93303.0,clean,
3 7301607386,knoxville,11590,2014.0,chevrolet,spark ev 2lt hatchback,good,,electric,61852.0,clean,other,KL
4 7301607388,beaumont / port arthur,11590,2014.0,chevrolet,spark ev 2lt hatchback,good,,electric,61852.0,c
5 7301607390,green bay,30990,2017.0,lincoln,mkx reserve sport,good,6 cylinders,gas,23545.0,clean,automatic
6 7301607395,toledo,29990,2017.0,lincoln,mkx reserve sport,good,6 cylinders,gas,23545.0,clean,automatic,2L
7 7301607396,flint,30990,2017.0,lincoln,mkx reserve sport,good,6 cylinders,gas,23545.0,clean,automatic,2LM
8 7301607402,cookeville,38990,2020.0,audi,q5 45 tfsi premium plus,good,,gas,19067.0,clean,other,WA1BNAFYXL
9 7301607482,rochester,5500,2011.0,,BLUE BIRD BLUE BIRD,,95023.0,,automatic,1BAKFC5A4CF287008,,,,ny
10 7301607519,albany,200,2005.0,,NONE,good,,other,111000.0,clean,other,,,,white,ny
11 7301607572,sierra vista,29990,2017.0,bmw,3-series,,,other,64677.0,clean,automatic,,rwd,,sedan,,az
12 7301608024,boone,5500,2006.0,,2006 SILVERADO 2500 HD,excellent,8 cylinders,gas,330500.0,clean,automatic,
13 7301608030,iowa city,700,1993.0,ford,ranger,fair,6 cylinders,gas,160000.0,clean,automatic,,4wd,,,green,i
14 7301608138,greenville / upstate,0,2017.0,chevrolet,colorado,good,4 cylinders,gas,27944.0,clean,automatic
15 7301608152,greenville / upstate,0,2019.0,dodge,grand caravan,good,,gas,96887.0,clean,other,2C4RDGBG3KR61
16 7301608251,richmond,4000,1994.0,mitsubishi,3000gt,good,6 cylinders,gas,149500.0,clean>manual,,fwd,,hatch
17 7301609037, chattanooga,12990,2011.0,acura,tsx sedan 4d,good,,other,85757.0,clean,other,JH4CU2F63BC006531
18 7301609038,altoona-johnstown,24590,2018.0,lincoln,mkz select sedan 4d,good,,gas,16081.0,clean,automatic,
19 7301609048,port huron,43990,2020.0,buick,enclave avenir sport,good,6 cylinders,gas,5217.0,clean,other,5C
20 7301609049,athens,38990,2017.0,acura,mdx sport hybrid sh-awd,good,,other,9326.0,clean,other,5FRYD7H59HB0
21 7301609055,macon / warner robins,34590,2020.0,infiniti,qx50 luxe sport utility,good,,other,3542.0,clean,
```

Data Transformation



Derived Columns & Handling Null Values

Column name *

CarInfo

Expression

```
@(id=id,
  price=iifNull(price,0,price),
  year=iifNull(year,2000,year),
  manufacturer=iifNull(manufacturer,"unknown",manufacturer),
  model=iifNull(model,"unknown",model),
  condition=iifNull(condition,"unknown",condition),
  cylinders=iifNull(cylinders,"unknown",cylinders),
  fuel=iifNull(fuel,"unknown",fuel),
  odometer=iifNull(odometer,"unknown",odometer),
  title_status=iifNull(title_status,"unknown",title_status),
  transmission=iifNull(transmission,"unknown",transmission),
  VIN=iifNull(VIN,"unknown",VIN),
  drive=iifNull(drive,"unknown",drive),
  size=iifNull(size,"unknown",size),
  type=iifNull(type,"unknown",type),
  paint_color=iifNull(paint_color,"unknown",paint_color))
```

Data Aggregation

Aggregate settings | Optimize | Inspect | Data preview ●

Output stream name * [Learn more](#)

Description [Reset](#)

Incoming stream *

Group by **Aggregates**

Columns	Name as	
<input type="text" value="abc state"/>	state	+
<input type="text" value="abc region"/>	region	+

Data in Blob Storage "Output"

```
1 {"state":"in","RegionInfo":[{"region":"terre haute","CarInfo":{"id":"7301609173","price":"11590","year":2019}}]}
2 {"state":"ct","RegionInfo":[{"region":"northwest CT","CarInfo":{"id":"7301611298","price":"41590","year":2019}}]}
3 {"state":"tx","RegionInfo":[{"region":"college station","CarInfo":{"id":"7301613420","price":"45590","year":2019}}]}
4 {"state":"al","RegionInfo":[{"region":"huntsville / decatur","CarInfo":{"id":"7301611362","price":"32590","year":2019}}]}
5 {"state":"ca","RegionInfo":[{"region":"visalia-tulare","CarInfo":{"id":"7301613331","price":"14590","year":2019}}]}
6 {"state":"ga","RegionInfo":[{"region":"macon / warner robins","CarInfo":{"id":"7301609055","price":"34590","year":2019}}]}
7 {"state":"ky","RegionInfo":[{"region":"lexington","CarInfo":{"id":"7301610754","price":"9942","year":2019}}]}
8 {"state":"mi","RegionInfo":[{"region":"holland","CarInfo":{"id":"7301611282","price":"29990","year":2019}}]}
9 {"state":"sc","RegionInfo":[{"region":"columbia","CarInfo":{"id":"7301609709","price":"3650","year":2019}}]}
10 {"state":"co","RegionInfo":[{"region":"fort collins / north CO","CarInfo":{"id":"7301613362","price":"32590","year":2019}}]}
11 {"state":"ny","RegionInfo":[{"region":"buffalo","CarInfo":{"id":"7301612667","price":"4500","year":2019}}]}
12 {"state":"oh","RegionInfo":[{"region":"tuscarawas co","CarInfo":{"id":"7301611306","price":"17590","year":2019}}]}
13 {"state":"pa","RegionInfo":[{"region":"york","CarInfo":{"id":"7301613441","price":"17590","year":2020}}]}
14 {"state":"va","RegionInfo":[{"region":"lynchburg","CarInfo":{"id":"7301612124","price":"6500","year":2019}}]}
15 {"state":"fl","RegionInfo":[{"region":"st augustine","CarInfo":{"id":"7301613345","price":"24590","year":2019}}]}
16 {"state":"nc","RegionInfo":[{"region":"fayetteville","CarInfo":{"id":"7301611333","price":"41590","year":2019}}]}
17 {"state":"ks","RegionInfo":[{"region":"topeka","CarInfo":{"id":"7301611328","price":"23590","year":2019}}]}
18 {"state":"az","RegionInfo":[{"region":"sierra vista","CarInfo":{"id":"7301607572","price":"29990","year":2019}}]}
19 {"state":"wi","RegionInfo":[{"region":"kenosha-racine","CarInfo":{"id":"7301609064","price":"38990","year":2019}}]}
20 {"state":"wv","RegionInfo":[{"region":"charleston","CarInfo":{"id":"7301609062","price":"24590","year":2019}}]}
21 {"state":"tn","RegionInfo":[{"region":"chattanooga","CarInfo":{"id":"7301609037","price":"12990","year":2019}}]}
22 {"state":"ok","RegionInfo":[{"region":"texoma","CarInfo":{"id":"7301609169","price":"41590","year":2019}}]}
```



 Preview

Data Pipeline



Type *

BlobEventsTrigger

Account selection method * ⓘ

☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ

Azure for Students (79914754-ab90-485d-a04f-0920cf3486a0) ▼

Storage account name * ⓘ

nancyblob ▼



Container name * ⓘ

input ▼

Blob path begins with ⓘ

used

Blob path ends with ⓘ

.csv

Event * ⓘ

☒ Blob created ☒ Blob deleted

Ignore empty blobs * ⓘ


☒ Yes ☐ No

Event Trigger

Document in Cosmos DB

[Home](#) [UsedCar\(Compl... x](#)

SELECT * FROM c [Edit Filter](#)

id	/s... 
1989d2...	wv
0a8e1e...	ga
95c05e...	ms
e79ce0...	ut
740fed...	ny
858206...	wy
1488a8...	or
3aaa58...	de
9f9b3b...	co
2a1896...	ca
60a191...	sc
ea88e0...	nv
f9d684...	tx
Load more	

1 {

2 "state": "wv",

3 "RegionInfo": [

4 {

5 "region": "southern WV",

6 "CarInfo": [

7 {

8 "price": 10000,

9 "year": 2002,

10 "manufacturer": "lincoln",

11 "model": "town car",

12 "condition": "excellent",

13 "cylinders": "8 cylinders",

14 "fuel": "gas",

15 "odometer": "249000.0",

16 "title_status": "clean",

17 "transmission": "automatic",

18 "VIN": "unknown",

19 "drive": "unknown",

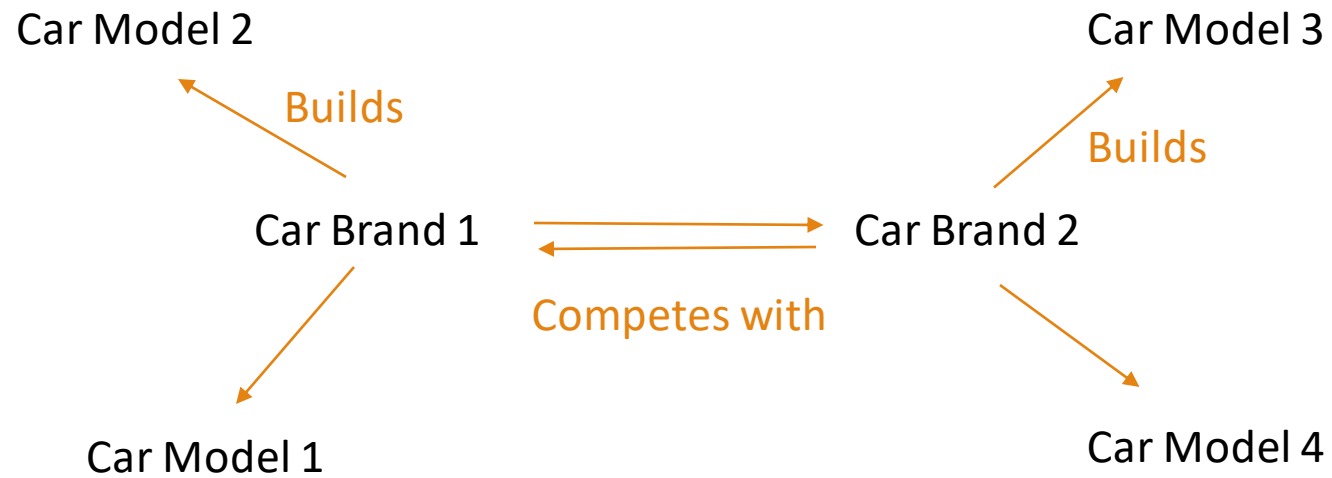
20 "size": "unknown",

21 "type": "unknown",

22 "paint_color": "blue"

23 },

Graph Implementation



Graph: Data Extraction

```
1 maker_model = pd.DataFrame(columns = ['Manufacturer', 'Model', 'Type'])
2
3 manufacturer = []
4 model = []
5 car_type = []
6
7
8 for j in range(used_car_test.shape[0]):
9     manufacturer.append(used_car_test.manufacturer[j])
10    model.append(used_car_test.model.str.split(" ")[j][0])
11    car_type.append(used_car_test.type[j])
12
13 maker_model['Manufacturer'] = manufacturer
14 maker_model['Model'] = model
15 maker_model['Type'] = car_type
```

```
1 maker_model.head()
```

	Manufacturer	Model	Type
0	gmc	sierra	pickup
1	chevrolet	silverado	pickup
2	chevrolet	silverado	pickup
3	toyota	tundra	pickup
4	ford	f-150	truck

Graph Implementation

```
def create_model_vertex(input_csv):
    model_vertex = []
    for i in maker[maker['Manufacturer'] == manu].Model.unique():
        v =
            "g.addV('model').property('id','{model}').property('Name','{model}').property('type','{m_t
type}').property('pk', 'pk')".format(model = i, m_type = maker.columns[1])
        model_vertex.append(v)
    return model_vertex
```

```
_gremlin_insert_vertices_maker = create_maker_vertex(maker)
_gremlin_insert_vertices_model = create_model_vertex(maker)

_gremlin_insert_vertices = _gremlin_insert_vertices_maker + _gremlin_insert_vertices_model
```

```
def insert_vertices(client):
    for query in _gremlin_insert_vertices:
        print("\n> {0}\n".format(query))
        callback = client.submitAsync(query)
        if callback.result() is not None:
            print("\tInserted this vertex:\n\t{0}".format(
                callback.result().all().result()))
        else:
            print("Something went wrong with this query: {0}".format(query))
        print("\n")
        print_status_attributes(callback.result())
        print("\n")
```

Graph Implementation

```
[project_graph.py]
Welcome to team 8 graph database!
Insert vertices into the graph database. Press any key to continue...

> g.addV('manufacturer').property('id', 'chevrolet').property('Name', 'chevrolet').property('type', 'manufacturer').property('pk', 'pk')

    Inserted this vertex:
    [{'id': 'chevrolet', 'label': 'manufacturer', 'type': 'vertex', 'properties':
{'Name': [{'id': '114f7e57-f055-4503-8124-6d9ccdca41e2', 'value': 'chevrolet'}], 'type':
': [{'id': '5bec5891-dc71-4c39-bb8e-ed5ff25679e9', 'value': 'manufacturer'}], 'pk': [{
'id': 'chevrolet|pk', 'value': 'pk'}]}}]
```


Graph Implementation

Results

ford

bmw

f350

e-450

f-350

f450

econoline

explorer

escape


ranger

fiesta

fusion

f-250

Graph



> ford

Properties

id

ford

label

manufacturer

Name

ford

type

Manufacturer

pk

pk

Sources

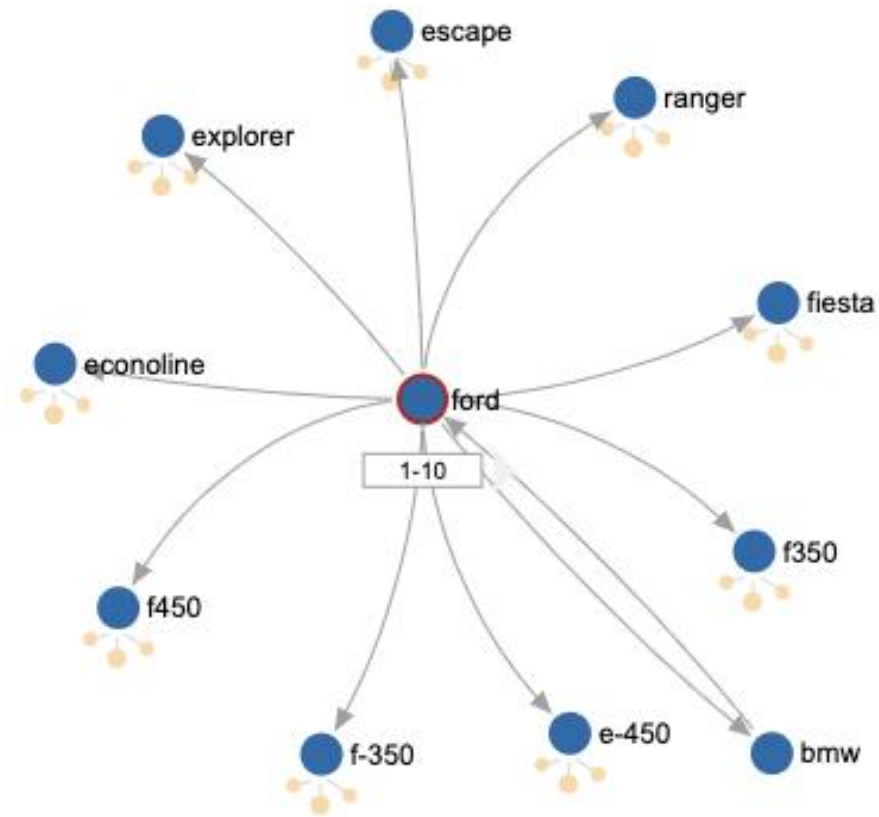
No sources found

Targets

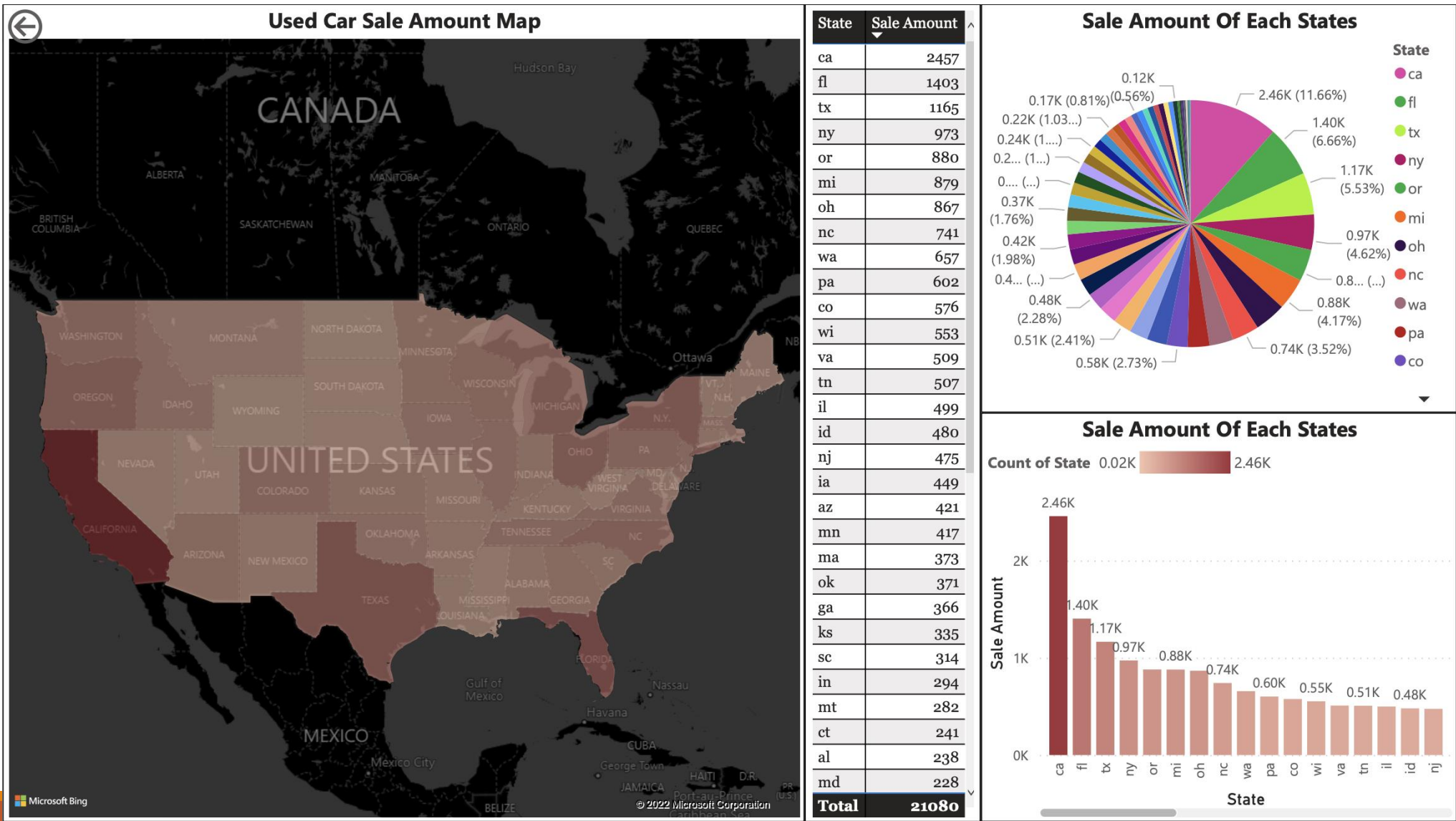
Target	Edge label
bmw	competes_with
f350	builds
e-450	builds
f-350	builds

17

Graph Implementation

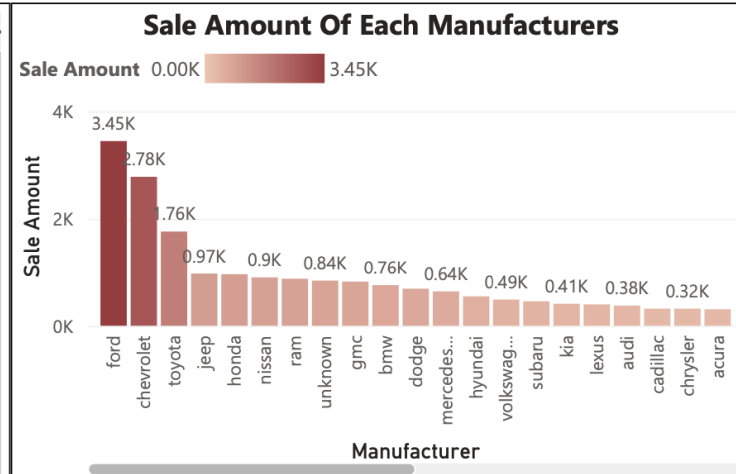


Used Car Sales By States

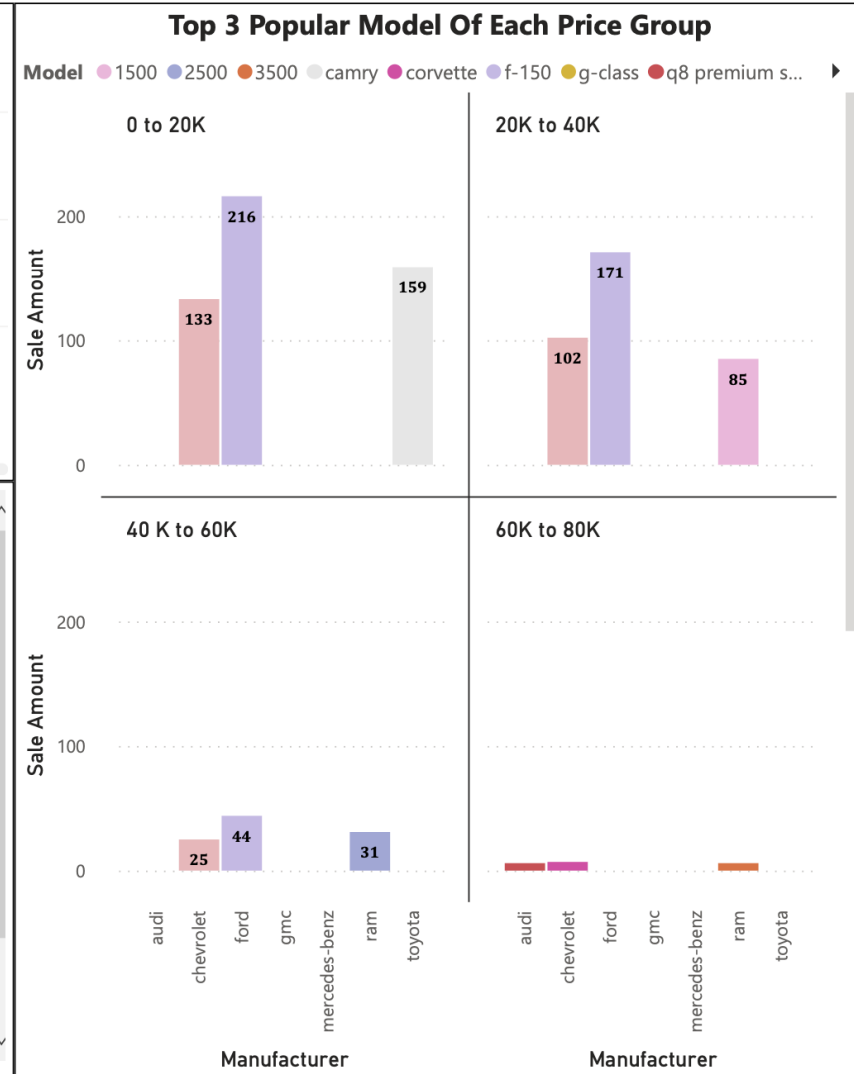


Manufacturer Analysis

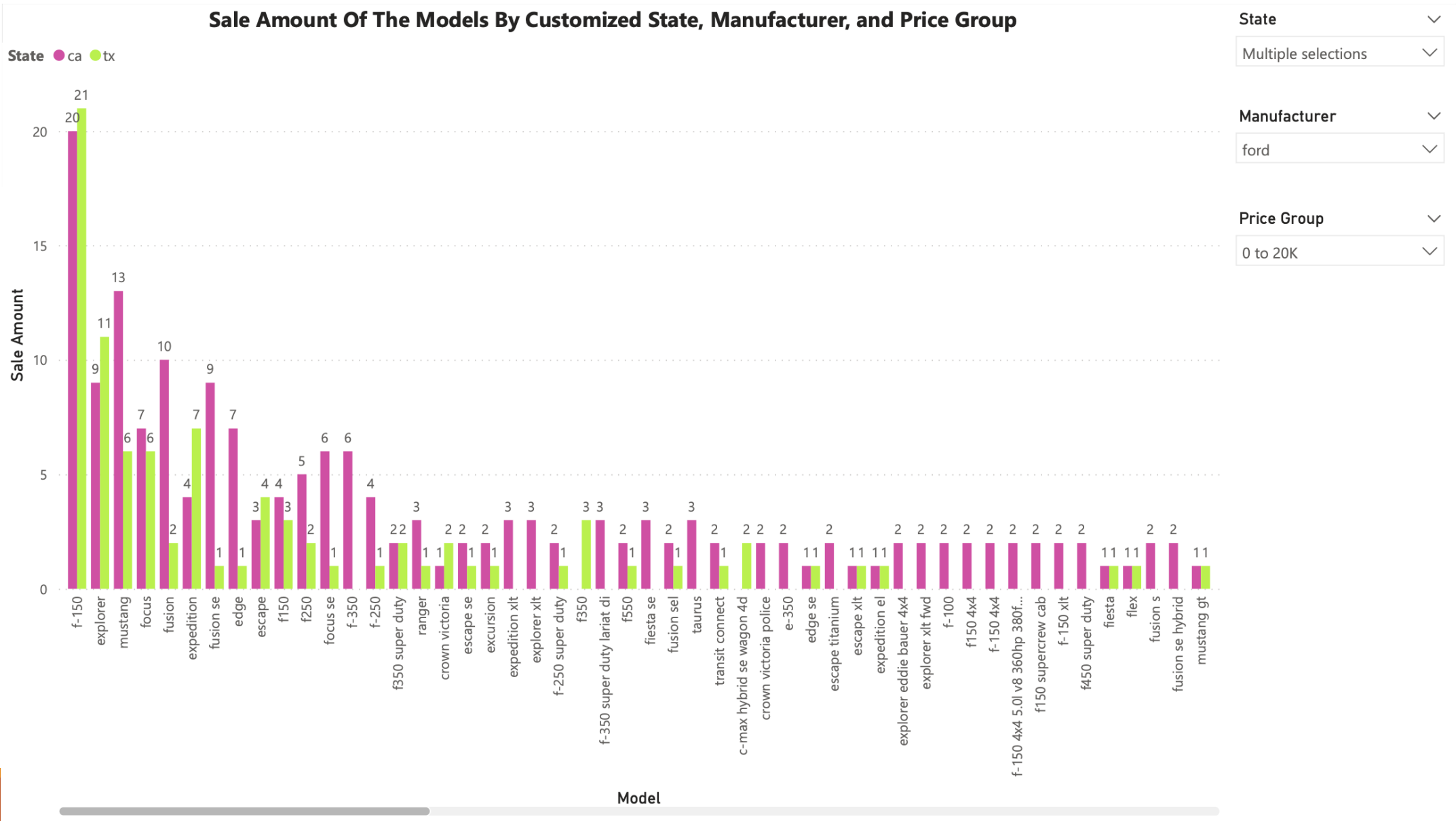
Manufacturer	Sale Amount
ford	3448
chevrolet	2783
toyota	1764
jeep	973
honda	960
nissan	900
ram	875
unknown	841
gmc	822
bmw	756
dodge	690
mercedes-benz	640
hyundai	547
volkswagen	488
subaru	455
kia	409
lexus	398
audi	375
cadillac	319
chrysler	319
acura	307
buick	273
infiniti	254
mazda	253
lincoln	199
mitsubishi	168
volvo	150
jaguar	108
mini	108
penton	86
Total	21080



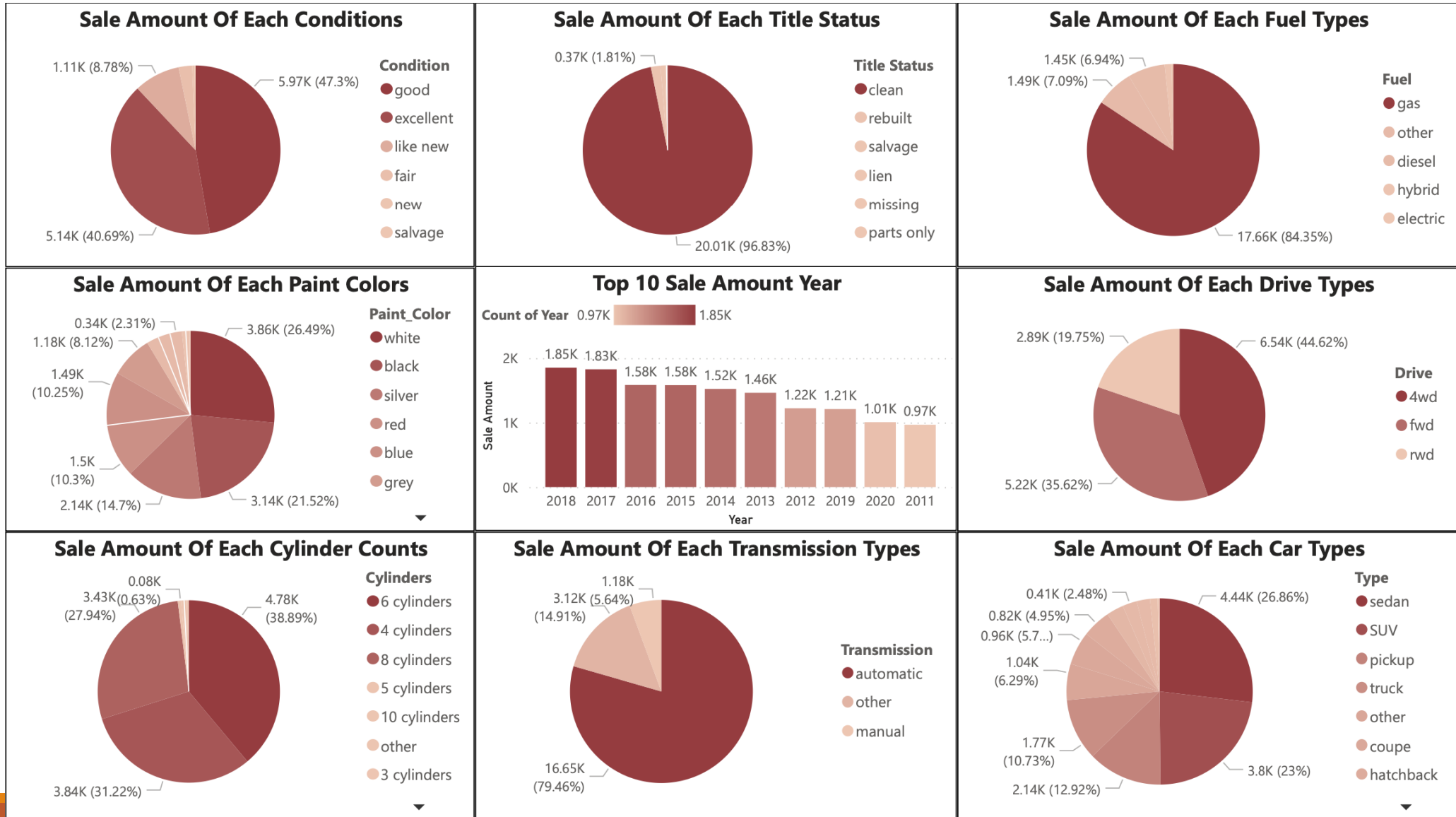
Price	Sale Amount	Manufacturer	Model
0 to 20K	133	chevrolet	silverado 1500
0 to 20K	216	ford	f-150
0 to 20K	159	toyota	camry
20K to 40K	102	chevrolet	silverado 1500
20K to 40K	171	ford	f-150
20K to 40K	85	ram	1500
40 K to 60K	25	chevrolet	silverado 1500
40 K to 60K	44	ford	f-150
40 K to 60K	31	ram	2500
60K to 80K	6	audi	q8 premium sport utility 4d
60K to 80K	7	chevrolet	corvette
60K to 80K	6	ram	3500
Above 80K	9	chevrolet	corvette
Above 80K	2	ford	super duty f-450 pickup
Above 80K	2	gmc	sierra 3500 lifted diesel 3500hd c
Total	1004		



Customized Sales Analysis



Sales By Car Attributes



Thank You!

