Business questions:

1. Describe the correlation of vehicles price on other factors. Find most significant.

2. Find the most selling region to go to

3. Find underestimated vehicles TO make the profitable deal

1. Calculating correlation koefficient for the manufacturers prices with year and odometer:

-- create view for the average data

**CREATE** **OR** **REPLACE** **VIEW** avg\_data **AS** -- I need average DATA FOR the correlations

**SELECT**

vt.manufacturer,

**round**(**avg**(p.price), 2) **AS** avg\_price,

**round**(**avg**(p.odometer)) **AS** avg\_odometer,

**avg**(p."year") **AS** avg\_year

**FROM** uv.publications p

**JOIN** uv.vehicle\_titles vt **USING** (vt\_id)

**GROUP** **BY** vt.manufacturer;

**SELECT** \* **FROM** avg\_data;

Result:

|  |  |  |  |
| --- | --- | --- | --- |
| manufacturer | avg\_price | avg\_odometer | avg\_year |
| acura | 24752.45 | 56454 | 2015.3935965334617236 |
| alfa-romeo | 29978.82 | 33723 | 2017.7956298200514139 |
| aston-martin | 56230.46 | 19826 | 2013.0000000000000000 |
| audi | 28182.45 | 55256 | 2015.4858646909728992 |
| bmw | 24122.87 | 64589 | 2014.0409157882806302 |
| buick | 19460.77 | 67093 | 2014.3441138421733506 |
| cadillac | 25987.35 | 62238 | 2014.4484038604305865 |
| chevrolet | 24198.97 | 83766 | 2012.9081182896967414 |
| chrysler | 13702.00 | 87342 | 2012.6294579172610556 |
| datsun | 19784.54 | 73105 | 1976.6153846153846154 |
| dodge | 20390.36 | 74389 | 2013.9649235720032180 |
| ferrari | 140656.76 | 22224 | 2007.0204081632653061 |
| fiat | 14406.56 | 45160 | 2015.5708737864077670 |
| ford | 25579.64 | 89741 | 2013.2431164479195755 |
| gmc | 29850.24 | 84281 | 2013.9012213413327619 |
| harley-davidson | 13460.58 | 40664 | 2008.7741935483870968 |
| honda | 14153.70 | 97585 | 2012.3609941872118661 |
| hyundai | 14133.30 | 73212 | 2014.7870695713281799 |
| infiniti | 24444.64 | 58995 | 2015.2444987775061125 |
| jaguar | 30978.75 | 34159 | 2015.8751773049645390 |
| jeep | 24062.84 | 71631 | 2014.0206925236560258 |
| kia | 14886.15 | 66390 | 2015.5336170212765957 |
| land rover | 5716.33 | 111601 | 2001.0000000000000000 |
| lexus | 23257.00 | 76014 | 2012.9739542225730071 |
| lincoln | 25015.90 | 55220 | 2014.9730856709628506 |
| mazda | 16839.54 | 71659 | 2013.9366621983914209 |
| mercedes-benz | 25912.97 | 67847 | 2012.8737298017657838 |
| mercury | 5917.09 | 131176 | 2003.3459459459459459 |
| mini | 18174.48 | 55195 | 2014.4588804422944022 |
| mitsubishi | 17883.35 | 54675 | 2015.4444444444444444 |
| nissan | 16347.46 | 79673 | 2014.2764325627259779 |
| pontiac | 9283.93 | 116654 | 2003.0610465116279070 |
| porsche | 37586.41 | 70905 | 2010.4444444444444444 |
| ram | 34187.70 | 79961 | 2015.3437925004857198 |
| rover | 30748.19 | 60022 | 2014.7984693877551020 |
| saturn | 7003.53 | 104821 | 2006.7394736842105263 |
| subaru | 16631.44 | 83390 | 2013.8386167146974063 |
| tesla | 39015.91 | 35197 | 2016.7363112391930836 |
| toyota | 20959.06 | 90076 | 2012.8408437200383509 |
| volkswagen | 16613.98 | 62219 | 2014.2731021555763824 |
| volvo | 24024.76 | 65072 | 2014.6418491484184915 |

-- 1.1 price with odometer (numerical - numerical) --- Pearson coefficient RESULT --- 0.4131681089025605

**SELECT** **sum**((p.price - avg\_data.avg\_price) \* (p.odometer - avg\_data.avg\_odometer))/ |/(**sum**((p.price - avg\_data.avg\_price)^2)\* **sum**((p.odometer - avg\_data.avg\_odometer)^2)) **AS** price\_odometer

**FROM** uv.publications p

**JOIN** uv.vehicle\_titles vt **USING**(vt\_id)

**JOIN** avg\_data **ON** avg\_data.manufacturer = vt.manufacturer;

-- 1.2. price with YEAR (numerical - numerical) --- Pearson coefficient RESULT 0.4752106053941027

**SELECT** **sum**((p.price - avg\_data.avg\_price) \* (p."year" - avg\_data.avg\_year))/ |/(**sum**((p.price - avg\_data.avg\_price)^2)\* **sum**((p."year"- avg\_data.avg\_year)^2)) **AS** price\_year

**FROM** uv.publications p

**JOIN** uv.vehicle\_titles vt **USING**(vt\_id)

**JOIN** avg\_data **ON** avg\_data.manufacturer = vt.manufacturer;

-- 1.3. price with year for models of the cars - Spearman coefficient

**SELECT** manufacturer,

1 - (6\***sum**(rank\_diff^2) / (**count**(manufacturer)\*(**count**(manufacturer)^2-1))) **AS** correlation\_spierman

**FROM**

(**SELECT** vt.manufacturer,

p.price,

p."year",

**RANK**() **OVER** (**PARTITION** **BY** vt.manufacturer **ORDER** **BY** p.price) **AS** price\_rank, -- ranking models by prices

**DENSE\_RANK** () **OVER** (**PARTITION** **BY** vt.manufacturer **ORDER** **BY** p."year" **DESC**) **AS** year\_rank,-- ranking models BY YEAR

**RANK**() **OVER** (**PARTITION** **BY** vt.manufacturer **ORDER** **BY** p.price) - **DENSE\_RANK** () **OVER** (**PARTITION** **BY** vt.manufacturer **ORDER** **BY** p."year" **desc**) **AS** rank\_diff

**FROM** uv.publications p

**JOIN** uv.vehicle\_titles vt **using**(vt\_id)

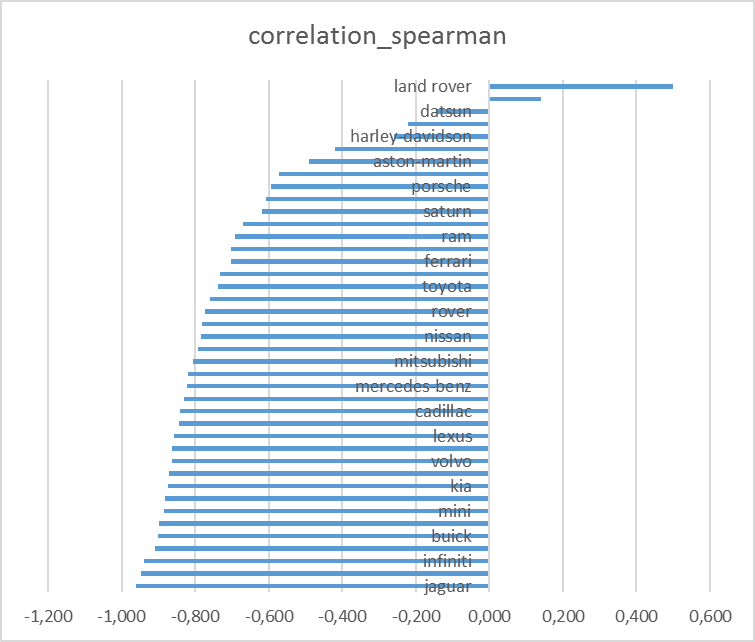
**GROUP** **BY** vt.manufacturer, p.price, p."year") tab

**GROUP** **BY** manufacturer

**ORDER** **BY** correlation\_spierman;

RESULT:

|  |  |
| --- | --- |
| manufacturer | correlation\_spearman |
| jaguar | -0,962 |
| lincoln | -0,948 |
| infiniti | -0,939 |
| acura | -0,910 |
| buick | -0,901 |
| subaru | -0,900 |
| mini | -0,884 |
| honda | -0,882 |
| kia | -0,874 |
| mazda | -0,872 |
| volvo | -0,863 |
| hyundai | -0,863 |
| lexus | -0,858 |
| audi | -0,843 |
| cadillac | -0,841 |
| chrysler | -0,830 |
| mercedes-benz | -0,822 |
| bmw | -0,818 |
| mitsubishi | -0,806 |
| volkswagen | -0,793 |
| nissan | -0,785 |
| dodge | -0,781 |
| rover | -0,772 |
| gmc | -0,760 |
| toyota | -0,738 |
| jeep | -0,733 |
| ferrari | -0,703 |
| alfa-romeo | -0,701 |
| ram | -0,690 |
| ford | -0,669 |
| saturn | -0,619 |
| fiat | -0,607 |
| porsche | -0,595 |
| chevrolet | -0,572 |
| aston-martin | -0,492 |
| tesla | -0,419 |
| harley-davidson | -0,260 |
| mercury | -0,220 |
| datsun | -0,145 |
| pontiac | 0,141 |
| land rover | 0,500 |



-- 2. dependence of number selling cars on region. Find the most selling region - where to go

**SELECT**

state,

region,

quantity\_of\_sales

**FROM**

(**SELECT** l.state,

l.region,

**count**(p.price) **AS** quantity\_of\_sales,

**RANK** () **OVER** (**PARTITION** **BY** l.state **ORDER** **BY** **count**(p.price) **DESC**) **AS** state\_rank

**FROM** uv.publications p

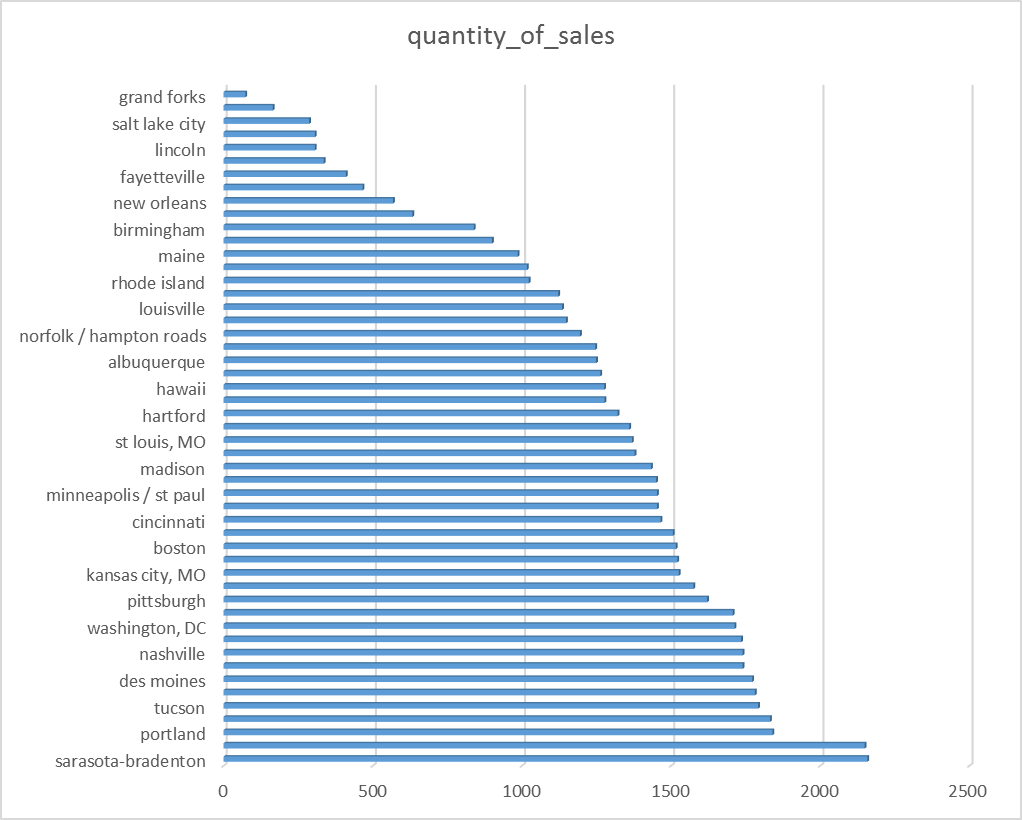
**JOIN** uv.locations l **ON** p.loc\_id = l.loc\_id

**GROUP** **BY** l.region, l.state) tab

**WHERE** state\_rank <= 3;

RESULT:

|  |  |  |
| --- | --- | --- |
| state | region | quantity\_of\_sales |
| fl | sarasota-bradenton | 2158 |
| ak | anchorage / mat-su | 2149 |
| or | portland | 1840 |
| ca | stockton | 1832 |
| az | tucson | 1792 |
| nj | central NJ | 1781 |
| ia | des moines | 1772 |
| id | boise | 1740 |
| tn | nashville | 1740 |
| mi | grand rapids | 1735 |
| dc | washington, DC | 1713 |
| wa | seattle-tacoma | 1707 |
| pa | pittsburgh | 1621 |
| md | baltimore | 1575 |
| ks | kansas city, MO | 1526 |
| ny | new york city | 1521 |
| ma | boston | 1516 |
| tx | san antonio | 1506 |
| oh | cincinnati | 1465 |
| co | colorado springs | 1453 |
| mn | minneapolis / st paul | 1453 |
| nh | new hampshire | 1450 |
| wi | madison | 1433 |
| ok | oklahoma city | 1378 |
| il | st louis, MO | 1369 |
| nv | las vegas | 1360 |
| ct | hartford | 1321 |
| vt | vermont | 1277 |
| hi | hawaii | 1276 |
| ga | atlanta | 1263 |
| nm | albuquerque | 1249 |
| in | indianapolis | 1246 |
| va | norfolk / hampton roads | 1195 |
| nc | raleigh / durham / CH | 1148 |
| ky | louisville | 1135 |
| mt | billings | 1122 |
| ri | rhode island | 1023 |
| sc | greenville / upstate | 1017 |
| me | maine | 986 |
| ct | new haven | 900 |
| al | birmingham | 839 |
| mo | columbia / jeff city | 633 |
| la | new orleans | 568 |
| de | delaware | 466 |
| ar | fayetteville | 410 |
| wv | charleston | 336 |
| ne | lincoln | 306 |
| sd | sioux falls / SE SD | 306 |
| ut | salt lake city | 287 |
| ms | north mississippi | 165 |
| nd | grand forks | 73 |



-- 3. underestimated vehicles TO make the profitable deal with the production year above 2000 having gas engine, automatic transmission and 4wd drive

**SELECT** url,

posting\_date,

region,

manufacturer,

model,

**to\_char**(price, '99 999.00') **AS** "price $",

**to\_char**(**round**(avg\_price\_model, 2), '99 999.00') **AS** "avg\_price $",

**to\_char**(avg\_price\_model - price, '99 999.99') **AS** "profit $"

**FROM**

(**SELECT** p.url,

p.posting\_date,

l.region,

vt.manufacturer,

vt.model,

p.price,

**avg**(p.price) **OVER** (**PARTITION** **BY** vt.model) **AS** avg\_price\_model

**FROM** uv.publications p

**JOIN** uv.locations l **USING** (loc\_id)

**JOIN** uv.vehicle\_titles vt **USING** (vt\_id)

**JOIN** uv.v\_engines ve **ON** ve.veng\_id = p.veng\_id

**JOIN** uv.vehicle\_body vb **USING** (vb\_id)

**JOIN** uv.drives d **ON** d.drive\_id = vb.drive\_id

**JOIN** uv.transmissions t **ON** t.trm\_id = vb.trm\_id

**WHERE** p."year" >= 2000

**AND** ve.fuel = 'gas'

**AND** d.drive\_type = '4wd'

**AND** t.trm\_type = 'automatic'

**GROUP** **BY** vt.manufacturer, vt.model, p.price, p.url, p.posting\_date, l.region) tab

**WHERE** avg\_price\_model - price > 0

**AND** price > 20000 **AND** price < 40000

**ORDER** **BY** avg\_price\_model - price **DESC**

**LIMIT** 10;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | url | posting\_date | region | manufacturer | model | price $ | avg\_price $ | profit $ | | https://poconos.craigslist.org/ctd/d/east-stroudsburg-2011-porsche-panamera/7317070385.html | 2021-05-04 22:08:23.000 | poconos | porsche | panamera turbo | 37 990.00 | 76 295.00 | 38 305.00 | | https://wichita.craigslist.org/ctd/d/wichita-2007-porsche-911-4s-carrera/7305834749.html | 2021-04-12 15:24:01.000 | wichita | porsche | 911 | 37 990.00 | 64 630.00 | 26 640.00 | | https://knoxville.craigslist.org/ctd/d/lenoir-city-2016-lincoln-navigator/7303335196.html | 2021-04-07 16:54:09.000 | knoxville | lincoln | navigator l | 24 999.00 | 50 698.43 | 25 699.43 | | https://boulder.craigslist.org/ctd/d/longmont-2016-land-rover-range-rover/7314340229.html | 2021-04-29 09:48:54.000 | boulder | rover | sport hs | 37 999.00 | 61 599.67 | 23 600.67 | | https://cnj.craigslist.org/ctd/d/miami-2020-jeep-gladiator-sport-4x4-4dr/7311619722.html | 2021-04-23 18:34:34.000 | central NJ | jeep | gladiator | 27 950.00 | 50 858.17 | 22 908.17 | | https://austin.craigslist.org/ctd/d/miami-2020-jeep-gladiator-sport-4x4-4dr/7311614249.html | 2021-04-23 17:22:55.000 | austin | jeep | gladiator | 27 950.00 | 50 858.17 | 22 908.17 | | https://bend.craigslist.org/ctd/d/portland-2001-toyota-land-cruiser-new/7311698854.html | 2021-04-23 19:03:33.000 | bend | toyota | land cruiser | 21 995.00 | 41 507.15 | 19 512.15 | | https://madison.craigslist.org/ctd/d/madison-2010-ford-150-svt-raptor/7314736669.html | 2021-04-30 07:30:41.000 | madison | ford | f-150 svt raptor | 21 995.00 | 40 700.50 | 18 705.50 | | https://missoula.craigslist.org/ctd/d/missoula-2016-lincoln-navigator-reserve/7312375976.html | 2021-04-25 10:42:34.000 | missoula | lincoln | navigator reserve | 29 995.00 | 48 371.00 | 18 376.00 | | https://sanantonio.craigslist.org/ctd/d/san-antonio-2014-ford-150-limited-4x4/7316211038.html | 2021-05-03 10:19:44.000 | san antonio | ford | f-150 limited 4x4 | 25 600.00 | 43 863.33 | 18 263.33 | |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |