In [1]:
```
import pandas as pd
import numpy as np

# NLTK libraries
from nltk.corpus import stopwords
from nltk import word_tokenize
from nltk.stem import WordNetLemmatizer
import nltk

# Visualization libraries
import plotly.graph_objects as go
import seaborn as sns
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from PIL import Image # for world cloud image

# Spacy for preprocessing
import spacy
from spacy.lang.en.stop_words import STOP_WORDS
nlp = spacy.load('en_core_web_sm')


# To change date to datetime
from datetime import datetime
import re

from collections import Counter
import string
import scipy.sparse

# Gensim libraries
from gensim import corpora
from gensim.models.ldamulticore import LdaMulticore
import pyLDAvis.gensim
from gensim.models import CoherenceModel
from gensim import matutils

# To show all the columns
pd.set_option('display.max_columns', 200)
pd.set_option('display.max_colwidth', 300)

# to pickle dataframe
import pickle

# Avoid warnings
import warnings
warnings.filterwarnings("ignore")


# Enable logging for gensim - optional but important
import logging
logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=logging.ERRO
```

In [2]:
```
#importing our dataset
import pandas as pd
brand=pd.read_csv("Dataset.csv")
brand.head()
```

Out[2]:

	Sr no.	Review_Date	Author_Name	Vehicle_Title	Review_Title	Review	Rating

	Sr no.	Review_Date	Author_Name	Vehicle_Title	Review_Title	Review	Rating
0	0	on 03/07/13 12:29 PM (PST)	deltasierra	2013 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Outstanding large family van	With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limited features, and the Mercede...	4.125
1	1	on 07/06/18 15:50 PM (PDT)	Daniel r	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Back ac suck	Rear ac blow to slow that my kid do not want to be in the back seat.	3.000
2	2	on 03/26/18 14:30 PM (PDT)	Bobbie D.	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	we love ours!	This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!	5.000
3	3	on 05/14/16 09:50 AM (PDT)	Joe Flash	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	My 2014 Nissan NVP SL review	I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We drove our van from Florida t...	5.000
4	4	on 10/21/15 21:37 PM (PDT)	Sam	2015 Nissan NV Passenger Van 3500 S 3dr Van (5.6L 8cyl 5A)	Not for a family	I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the road well* great w...	3.000

DATA CLEANING

In [3]:
```
# Drop the Sr no. column and Author _name column
brand.drop(['Author_Name','Sr no.'],axis=1,inplace=True)
# CHeck the data info
brand.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11715 entries, 0 to 11714
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Review_Date    11715 non-null  object 
 1   Vehicle_Title  11715 non-null  object 
 2   Review_Title   11714 non-null  object 
 3   Review         11715 non-null  object 
 4   Rating         11715 non-null  float64
dtypes: float64(1), object(4)
memory usage: 457.7+ KB
```

In [4]:
```
# Check for nun values
brand.isnull().sum()
```

Out[4]:
```
Review_Date      0
Vehicle_Title    0
```

```
Review_Title     1
Review           0
Rating           0
dtype: int64
```

In [5]:
```
# use interpolate to get the nearest rating score
brand['Rating'] = brand['Rating'].interpolate()
```

In [6]:
```
# Join the Review_Title and Review columns
brand["review"] = brand["Review_Title"].map(str) + brand["Review"]
```

In [7]:
```
# droping the rows with nun review
brand.dropna(axis=0, how='all', thresh=None, subset=['Review'], inplace=True)
brand.head()
```

Out[7]:

	Review_Date	Vehicle_Title	Review_Title	Review	Rating	review
0	on 03/07/13 12:29 PM (PST)	2013 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Outstanding large family van	With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limited features, and the Mercede...	4.125	Outstanding large family van With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limit...
1	on 07/06/18 15:50 PM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Back ac suck	Rear ac blow to slow that my kid do not want to be in the back seat.	3.000	Back ac suck Rear ac blow to slow that my kid do not want to be in the back seat.
2	on 03/26/18 14:30 PM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	we love ours!	This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!	5.000	we love ours! This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!
3	on 05/14/16 09:50 AM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	My 2014 Nissan NVP SL review	I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We drove our van from Florida t...	5.000	My 2014 Nissan NVP SL review I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We ...
4	on 10/21/15 21:37 PM (PDT)	2015 Nissan NV Passenger Van 3500 S 3dr Van (5.6L 8cyl 5A)	Not for a family	I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the road well* great w...	3.000	Not for a family I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the ro...

In [8]:
```
# Check if we still have nun values
brand.isnull().sum()
```

Out[8]:
```
Review_Date      0
Vehicle_Title    0
Review_Title     1
Review           0
Rating           0
review           0
dtype: int64
```

In [9]:
```
# check the data type
brand.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11715 entries, 0 to 11714
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Review_Date    11715 non-null  object 
 1   Vehicle_Title  11715 non-null  object 
 2   Review_Title   11714 non-null  object 
 3   Review         11715 non-null  object 
 4   Rating         11715 non-null  float64
 5   review         11715 non-null  object 
dtypes: float64(1), object(5)
memory usage: 640.7+ KB
```

In [10]:
```
# spliting the Vehicle_title into year, car  name and model column
brand['year'] = brand.Vehicle_Title.str.split(' ').apply(lambda x:x[0])
brand['car_name'] = brand.Vehicle_Title.str.split(' ').apply(lambda x:x[1])
brand['model'] = brand.Vehicle_Title.str.split(' ').apply(lambda x:x[2])
brand
```

Out[10]:

	Review_Date	Vehicle_Title	Review_Title	Review	Rating	review	year	car_name	mode
0	on 03/07/13 12:29 PM (PST)	2013 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Outstanding large family van	With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limited features, and the Mercede...	4.125	Outstanding large family van With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limit...	2013	Nissan	NV
1	on 07/06/18 15:50 PM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	Back ac suck	Rear ac blow to slow that my kid do not want to be in the back seat.	3.000	Back ac suck Rear ac blow to slow that my kid do not want to be in the back seat.	2015	Nissan	NV

	Review_Date	Vehicle_Title	Review_Title	Review	Rating	review	year	car_name	mode
2	on 03/26/18 14:30 PM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	we love ours!	This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!	5.000	we love ours! This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!	2015	Nissan	NV
3	on 05/14/16 09:50 AM (PDT)	2015 Nissan NV Passenger Van 3500 SL 3dr Van (5.6L 8cyl 5A)	My 2014 Nissan NVP SL review	I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We drove our van from Florida t...	5.000	My 2014 Nissan NVP SL review I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We ...	2015	Nissan	NV
4	on 10/21/15 21:37 PM (PDT)	2015 Nissan NV Passenger Van 3500 S 3dr Van (5.6L 8cyl 5A)	Not for a family	I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the road well* great w...	3.000	Not for a family I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the ro...	2015	Nissan	NV
...	...	...	...	...	...	...	...	...	.

	Review_Date	Vehicle_Title	Review_Title	Review	Rating	review	year	car_name	mode
11710	on 07/22/17 17:08 PM (PDT)	2017 Nissan Versa Sedan 1.6 SV 4dr Sedan (1.6L 4cyl CVT)	Get car to build credit	I got my 2017 Nissan Versa back in March of this year, my biggest complaint is about the dealership that sold me the car. But I am not reviewing this major Dallas dealership, I am reviewing the car. I had been renting cars a lot, one of the rentals I drove was a Versa. I loved driving the renta...	4.000	Get car to build credit I got my 2017 Nissan Versa back in March of this year, my biggest complaint is about the dealership that sold me the car. But I am not reviewing this major Dallas dealership, I am reviewing the car. I had been renting cars a lot, one of the rentals I drove was a Versa. I ...	2017	Nissan	Vers
11711	on 06/27/17 15:59 PM (PDT)	2017 Nissan Versa Sedan 1.6 S 4dr Sedan (1.6L 4cyl 5M)	good for the elderly	We liked the car very much. However the dealer did not want to honor the certificate at San Bernardino Nissan because we wern't financing. We did eventually get the car because Pete arranged it.	4.000	good for the elderly We liked the car very much. However the dealer did not want to honor the certificate at San Bernardino Nissan because we wern't financing. We did eventually get the car because Pete arranged it.	2017	Nissan	Vers
11712	on 03/12/17 09:21 AM (PDT)	2017 Nissan Versa Sedan 1.6 S 4dr Sedan (1.6L 4cyl 5M)	Warranted if not driven outside	AC condenser not covered if (and apparently common) damaged from rock from road.Protected underneath,but not through front.Which is amazingly open and unprotected (clearly a design flaw.	3.000	Warranted if not driven outside AC condenser not covered if (and apparently common) damaged from rock from road.Protected underneath,but not through front.Which is amazingly open and unprotected (clearly a design flaw.	2017	Nissan	Vers
11713	on 03/08/17 09:23 AM (PST)	2017 Nissan Versa Sedan 1.6 S 4dr Sedan (1.6L 4cyl 5M)	reliable transportation	this is not a sports car but it is reliable transportation for a low price. handles winter weather conditions and city driving very well	5.000	reliable transportation this is not a sports car but it is reliable transportation for a low price. handles winter weather conditions and city driving very well	2017	Nissan	Vers

	Review_Date	Vehicle_Title	Review_Title	Review	Rating	review	year	car_name	mode
11714	on 12/09/16 12:31 PM (PST)	2017 Nissan Versa Sedan 1.6 S Plus 4dr Sedan (1.6L 4cyl CVT)	2017 Versa My First Car	I went to Nissan looking to see what kind of car I could afford. I was not expecting to walk out and get into a new car! However, the car was suggested to me since it would be my first car. I have learned a lot about the Versa since I have had it for about 2 1/2 months. The car is comfortable ...	4.000	2017 Versa My First Car I went to Nissan looking to see what kind of car I could afford. I was not expecting to walk out and get into a new car! However, the car was suggested to me since it would be my first car. I have learned a lot about the Versa since I have had it for about 2 1/2 months. ...	2017	Nissan	Vers

11715 rows × 9 columns

In [11]:
```
# taking only the date out from the Review_date column and putting in in a new column (dat
brand['date'] = brand['Review_Date'].str.extract(r"(\d{1,2}[/. ](?:\d{1,2}|January|Jan)[/.

# Change the date column to datetime
brand['date'] =  pd.to_datetime(brand['date'],format='%m/%d/%y')
```

In [12]:
```
# Drop all the unwanted columns
brand.drop(['Review_Date','Vehicle_Title','Review_Title','Review'],axis=1,inplace=True)
```

In [13]:
```
# Converting rating to int
brand['Rating'] = brand['Rating'].astype(int)

display(brand.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11715 entries, 0 to 11714
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Rating    11715 non-null  int32
 1   review    11715 non-null  object
 2   year      11715 non-null  object
 3   car_name  11715 non-null  object
 4   model     11715 non-null  object
 5   date      11715 non-null  datetime64[ns]
dtypes: datetime64[ns](1), int32(1), object(4)
memory usage: 594.9+ KB
None
```

In [14]:
```
# Extract review year, month and day name
brand['review_year'] = brand.date.dt.year
brand['month'] = brand.date.dt.month
brand['day'] = brand.date.dt.day
```

In [15]:
```
brand.head()
```

Out[15]:

	Rating	review	year	car_name	model	date	review_year	month	day

	Rating	review	year	car_name	model	date	review_year	month	day
0	4	Outstanding large family van With the expected arrival of our 6th child, our Toyota Sienna minivan was going to be too small for our needs. The thought of diving a huge 12-passenger van did not appeal to us. The choices for a long time have pretty much been either Ford or Chevy, which have limit...	2013	Nissan	NV	2013-03-07	2013	3	7
1	3	Back ac suck Rear ac blow to slow that my kid do not want to be in the back seat.	2015	Nissan	NV	2018-07-06	2018	7	6
2	5	we love ours! This is not a small astro van type.You will need to navigate certain parking lots,spaces and drive fast food drive thrus .This is why we bought it because it is extra roomey!	2015	Nissan	NV	2018-03-26	2018	3	26
3	5	My 2014 Nissan NVP SL review I am very satisfied with my 2014 Nissan NV SL. I use this van for my business deliveries and personal use. Camping, road trips, etc. We dont have any children so I store most of the seats in my warehouse. I wanted the passenger van for the rear air conditioning. We ...	2015	Nissan	NV	2016-05-14	2016	5	14
4	3	Not for a family I went from a Honda Odyssey to this van since our family grew and we needed more room. we have all the kids in boosters or car seats. Here are the pros:* smooth drive*comfortable driver configurations (however the shift stick gets in the way of adjusting the ac.)* handles the ro...	2015	Nissan	NV	2015-10-21	2015	10	21

PICKLE THE DATAFRAME

In [16]:
```
# Let's pickle it for later use
brand.to_pickle("brand_with_part_of_year.pkl")
```

In [17]:
```
# To see the percentage of each brands review in the dataset
brand_review_pct = brand['model'].value_counts(normalize = True) * 100
brand_review_pct
```

Out[17]:
```
Altima        14.007682
Sentra        12.121212
Frontier      11.259070
Maxima        10.004268
Pathfinder     9.312847
Titan          6.350832
Xterra         6.222791
Rogue          6.086214
Versa          5.727700
Z350           4.541187
Quest          4.430218
Armada         3.363210
Juke           1.442595
Z370           1.160905
Leaf           1.118224
Cube           1.067008
SX200          0.418267
Truck          0.401195
GTR            0.315834
NV200          0.196329
Murano         0.170721
SX240          0.128041
NV             0.128041
```

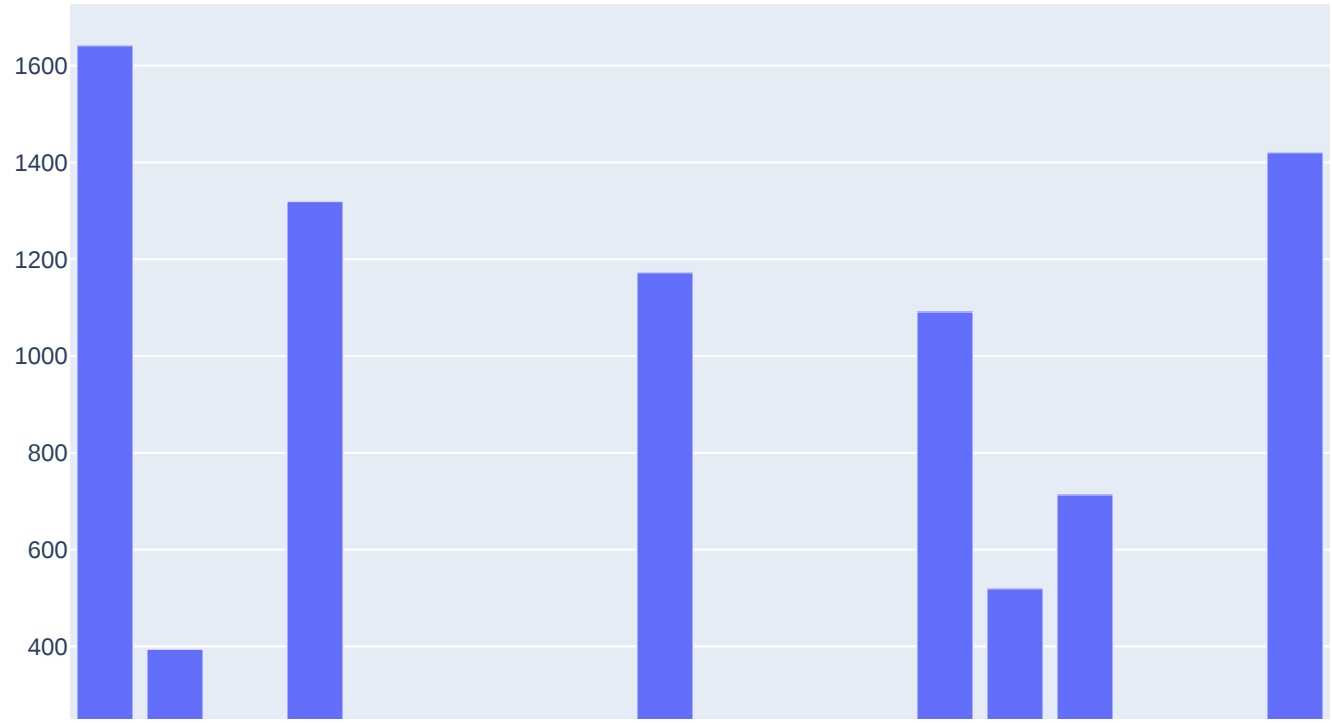
```
Kicks           0.025608
Name: model, dtype: float64
```

In [18]:
```
# create a Dataframe for the count of reviews of each brand
brand_review_count = brand.groupby('model').count()['review'].reset_index()
brand_review_count
```

Out[18]:

	model	review
0	Altima	1641
1	Armada	394
2	Cube	125
3	Frontier	1319
4	GTR	37
5	Juke	169
6	Kicks	3
7	Leaf	131
8	Maxima	1172
9	Murano	20
10	NV	15
11	NV200	23
12	Pathfinder	1091
13	Quest	519
14	Rogue	713
15	SX200	49
16	SX240	15
17	Sentra	1420
18	Titan	744
19	Truck	47
20	Versa	671
21	Xterra	729
22	Z350	532
23	Z370	136

In [19]:
```
# Using plotly to create Barchat
bar_go = go.Bar(x = brand_review_count['model'], y = brand_review_count['review'], name='R
fig = go.Figure(
    data=[bar_go],
    layout=go.Layout(width=1000, height=600, title='Brand Review Count', xaxis_title='make
fig.show()
```



In [20]:
```
# the count of each brand according to their rating
grouped_brand = brand.groupby([brand.model, brand.Rating]).size().reset_index().rename(col
grouped_brand
```

Out[20]:

	model	Rating	counts
0	Altima	1	84
1	Altima	2	124
2	Altima	3	229
3	Altima	4	752
4	Altima	5	452
...	...	...	...
103	Z370	1	2
104	Z370	2	3
105	Z370	3	8
106	Z370	4	66
107	Z370	5	57

108 rows × 3 columns

In [21]:
```
# Remove the numbers from the review
brand['review'] = brand['review'].apply(lambda x: re.sub(r'[^A-Za-z\s]', '', x))

# Convert the reviews to lowercase
brand['review'] = brand['review'].map(lambda x: x.lower())
brand.review
```

Out[21]:
```
0        outstanding large family van with the expected arrival of our th child our toyota
sienna minivan was going to be too small for our needs the thought of diving a huge passen
```

```
ger van did not appeal to us the choices for a long time have pretty much been either ford
or chevy which have limited featu...
1
back ac suck rear ac blow to slow that my kid do not want to be in the back seat
2
we love ours this is not a small astro van typeyou will need to navigate certain parking l
otsspaces and drive fast food drive thrus this is why we bought it because it is extra roo
mey
3        my  nissan nvp sl review i am very satisfied with my   nissan nv sl i use this va
n for my business deliveries and personal use camping road trips etc we dont have any chil
dren so i store most of the seats in my warehouse i wanted the passenger van for the rear
air conditioning we drove our van f...
4        not for a family i went from a honda odyssey to this van since our family grew an
d we needed more room we have all the kids in boosters or car seats here are the pros smoo
th drivecomfortable driver configurations however the shift stick gets in the way of adjus
ting the ac handles the road well g...

...
11710    get car to build credit i got my  nissan versa back in march of this year my bigg
est complaint is about the dealership that sold me the car but i am not reviewing this maj
or dallas dealership i am reviewing the car i had been renting cars a lot one of the renta
ls i drove was a versa i loved driv...
11711
good for the elderly we liked the car very much however the dealer did not want to honor t
he certificate at san bernardino nissan because we wernt financing we did eventually get t
he car because pete arranged it
11712
warranted if not driven outside ac condenser not covered if and apparently common damaged
from rock from roadprotected underneathbut not through frontwhich is amazingly open and un
protected clearly a design flaw
11713
reliable transportation this is not a sports car but it is reliable transportation for a l
ow price handles winter weather conditions and city driving very well
11714     versa my first car i went to nissan looking  to see what kind of car i could aff
ord i was not expecting to walk out and get into a new car however the car was suggested t
o me since it would be my first car i have learned a lot about the versa since i have had
it for about   months the car is co...
Name: review, Length: 11715, dtype: object
```

In [22]:
```
stop_words = stopwords.words('english')
# stop_words.extend([''])
```

In [23]:
```
def lematized_review(text): # text
    rev_text = nlp(text)
    # Extract lematized words in lower case format if not digits, not punctuation, not sto
    rev_text = ([token.lemma_.lower() for token in rev_text if not token.is_stop and token
    return rev_text
```

In [24]:
```
%%time
# Applying the function on the reviews

brand['review'] = brand['review'].apply(lematized_review)
```

```
Wall time: 8min 30s
```

In [25]:
```
#  Let's pickle it for later use
clean_brand_review = brand['review']
```

In [26]:
```
clean_brand_review
```

Out[26]:
```
0        [outstanding, large, family, expect, arrival, child, toyota, sienna, minivan, go,
small, need, thought, dive, huge, passenger, appeal, choice, long, time, pretty, ford, che
vy, limit, feature, mercedes, nice, expensive, pleased, learn, nissan, start, sell, passen
ger, van, price, ford, chevy, van,...
1
[suck, rear, blow, slow, want, seat]
2
```

```
[love, small, astro, typeyou, need, navigate, certain, parking, lotsspace, drive, fast, fo
od, drive, thrus, buy, extra, roomey]
3        [nissan, review, satisfied, nissan, business, delivery, personal, camping, road,
trip, child, store, seat, warehouse, want, passenger, rear, conditioning, drive, florida,
california, cross, country, trip, average, drive, rain, comfortable, stable, vehicle, niss
an, titan, engine, mile, engine, te...
4        [family, go, honda, odyssey, family, grow, need, room, kid, booster, seat, pro, s
mooth, drivecomfortable, driver, configuration, shift, stick, get, adjust, handle, road, g
reat, warranty, people, lanecon, wide, turn, difficult, maneuver, shopping, parking, lot,
carpool, lane, school, power, long,...

...
11710    [build, credit, nissan, versa, march, year, big, complaint, dealership, sell, rev
iew, major, dallas, dealership, review, rent, car, rental, drive, versa, love, drive, rent
al, city, average, mile, go, shortly, buy, buy, awesome, talk, handle, great, response, ti
ght, steer, great, acceleration, gr...
11711
[good, elderly, like, dealer, want, honor, certificate, bernardino, nissan, wernt, financi
ng, eventually, pete, arrange]
11712
[warrant, drive, outside, condenser, cover, apparently, common, damage, rock, roadprotect
e, underneathbut, frontwhich, amazingly, open, unprotecte, clearly, design, flaw]
11713
[reliable, transportation, sport, reliable, transportation, price, handle, winter, weathe
r, condition, city, drive]
11714    [versa, go, nissan, look, kind, afford, expect, walk, suggest, learn, versa, mont
h, comfortable, storage, space, roomy, especially, seat, engine, noisy, run, smoothly, pro
blem, manuevere, traffic, parking, spot, understand, subcompact, vehicle, feel, little, ti
ght, tall, person, drive, comfortab...
Name: review, Length: 11715, dtype: object
```

In [27]:

```
%%time
# Create Dictionary
id2word_1 = corpora.Dictionary(clean_brand_review)

# Create Corpus: Term Document Frequency
corpus_1 = [id2word_1.doc2bow(review) for review in clean_brand_review]

 # Build LDA model
ldamodel = LdaMulticore(corpus= corpus_1, num_topics =14, id2word=id2word_1,chunksize=2000
```

```
Wall time: 5min 45s
```

In [28]:

```
from pprint import pprint

pprint(ldamodel.show_topics(formatted=False))
```

```
[(4,
  [('nissan', 0.011548695),
   ('year', 0.008750253),
   ('need', 0.006348358),
   ('look', 0.0058920784),
   ('clutch', 0.005702093),
   ('purchase', 0.005324926),
   ('like', 0.00521279),
   ('content', 0.0050942176),
   ('go', 0.004985895),
   ('truck', 0.0041140765)]),
 (5,
  [('drive', 0.033866353),
   ('love', 0.02429238),
   ('like', 0.018760689),
   ('look', 0.016170982),
   ('nissan', 0.01238535),
   ('good', 0.011371653),
   ('quest', 0.010736363),
   ('great', 0.010419483),
   ('vehicle', 0.009565061),
   ('want', 0.008161708)]),
```

```
(2,
 [('great', 0.03301771),
  ('drive', 0.029159267),
  ('love', 0.02054576),
  ('good', 0.018593533),
  ('look', 0.013898792),
  ('altima', 0.0126085505),
  ('like', 0.011833748),
  ('nissan', 0.011615427),
  ('ride', 0.010343245),
  ('comfortable', 0.008913241)]),
(8,
 [('head', 0.0120985685),
  ('engine', 0.011371707),
  ('gasket', 0.010024597),
  ('like', 0.007090146),
  ('look', 0.0052630682),
  ('good', 0.005208821),
  ('thing', 0.003742452),
  ('go', 0.0037289076),
  ('build', 0.0033878905),
  ('car', 0.0032807007)]),
(9,
 [('juke', 0.01677488),
  ('love', 0.012130235),
  ('problem', 0.011957582),
  ('vehicle', 0.011124173),
  ('year', 0.010497515),
  ('sentra', 0.008691648),
  ('drive', 0.0077974973),
  ('buy', 0.007145621),
  ('nissan', 0.006908007),
  ('time', 0.006298915)]),
(11,
 [('nissan', 0.043709226),
  ('problem', 0.02385369),
  ('transmission', 0.022910547),
  ('mile', 0.017520772),
  ('dealer', 0.014209042),
  ('time', 0.013243434),
  ('replace', 0.012693162),
  ('go', 0.012510446),
  ('warranty', 0.011233813),
  ('issue', 0.011226975)]),
(12,
 [('mile', 0.04082909),
  ('year', 0.024665061),
  ('replace', 0.02344818),
  ('tire', 0.017638277),
  ('drive', 0.017119717),
  ('buy', 0.0171039),
  ('problem', 0.016083557),
  ('good', 0.015542733),
  ('great', 0.012860567),
  ('brake', 0.012006155)]),
(6,
 [('wheel', 0.0105341915),
  ('rear', 0.008187961),
  ('turn', 0.007876257),
  ('work', 0.007217642),
  ('good', 0.0067098853),
  ('radius', 0.00571174),
  ('roof', 0.0056933914),
  ('rack', 0.0055883112),
  ('driver', 0.005349992),
  ('truck', 0.00486692)]),
(7,
 [('mile', 0.029478924),
  ('drive', 0.022144772),
  ('mileage', 0.01698712),
```

```
   ('trip', 0.013855545),
   ('great', 0.013836557),
   ('highway', 0.013509349),
   ('average', 0.011897663),
   ('vehicle', 0.011732138),
   ('long', 0.009618385),
   ('city', 0.009248999)]),
 (0,
  [('seat', 0.024881423),
   ('good', 0.011179075),
   ('like', 0.010573516),
   ('rear', 0.009943004),
   ('interior', 0.0098110605),
   ('nice', 0.009218348),
   ('door', 0.008797866),
   ('great', 0.007482797),
   ('drive', 0.00733459),
   ('turn', 0.007146991)])]
```

In [29]:
```
# Compute Perplexity

#It's a measure of how good the model is. The lower the better. Perplexity is a negative v
print('\nPerplexity: ', ldamodel.log_perplexity(corpus_1))

# Compute Coherence Score
coherence_model_lda = CoherenceModel(model=ldamodel, texts=clean_brand_review, dictionary=
coherence_lda = coherence_model_lda.get_coherence()
print('\n Basic Ldamodel Coherence Score: ', coherence_lda)
```

```
Perplexity:  -7.352601011492008

 Basic Ldamodel Coherence Score:  0.36914298514524535
```

perplexity is a measurement of how well a probability distribution or probability model predicts a sample. It may be used to compare probability models. A low perplexity indicates the probability distribution is good at predicting the sample.

The coherence score is used in assessing the quality of the learned topics, the closer to 1 the better

In [30]:
```
    # Import mallet packages
import os
from gensim.models.wrappers import LdaMallet
from gensim.models.wrappers.ldamallet import malletmodel2ldamodel
os.environ.update({'MALLET_HOME': r'C:/Users/riya2/mallet/'})
# os.environ['MALLET_HOME'] = 'C:/Users/riya2/mallet/' # My mallet path, it is needed to i
```

In [31]:
```
%%time
import gensim
#  point the path to the mallet path on my computer
mallet_path = 'C:/Users/riya2/mallet/bin/mallet' #insert the path

NUM_TOPICS = 16
ldamallet = gensim.models.ldamodel.LdaModel(corpus_1, num_topics = NUM_TOPICS, id2word=id2
ldamallet.save('model5.gensim')
topics = ldamallet.print_topics(num_words=10)
for topic in topics:
    print(topic)
#Compute Coherence Score
coherence_model_ldamallet = CoherenceModel(model=ldamallet, texts=clean_brand_review, dict
coherence_ldamallet = coherence_model_ldamallet.get_coherence()
print('\n Mallet Coherence Score: ', coherence_ldamallet)
```

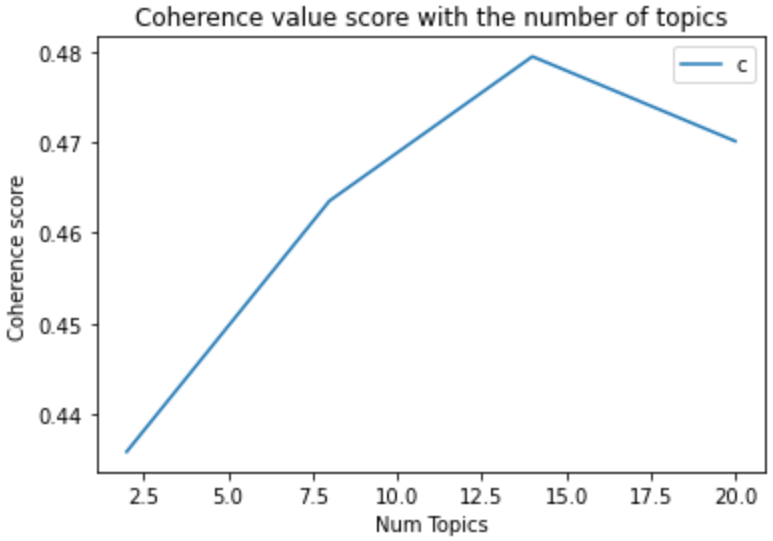
```
(0, '0.043*"noise" + 0.030*"seat" + 0.026*"rattle" + 0.026*"door" + 0.022*"window" + 0.018
*"driver" + 0.013*"rear" + 0.012*"turn" + 0.012*"passenger" + 0.012*"loud"')
(1, '0.117*"nissan" + 0.040*"transmission" + 0.026*"problem" + 0.021*"service" + 0.020*"wa
rranty" + 0.020*"vehicle" + 0.014*"dealer" + 0.012*"dealership" + 0.012*"customer" + 0.012
```

```
*"tell"')
(2, '0.034*"roadster" + 0.021*"dollar" + 0.017*"corvette" + 0.017*"spray" + 0.014*"particu
larly" + 0.014*"workhorse" + 0.014*"awsome" + 0.012*"chair" + 0.010*"bucket" + 0.010*"vers
us"')
(3, '0.030*"replace" + 0.027*"mile" + 0.025*"problem" + 0.019*"nissan" + 0.017*"go" + 0.01
5*"time" + 0.015*"buy" + 0.014*"engine" + 0.014*"brake" + 0.014*"transmission"')
(4, '0.039*"tundra" + 0.023*"diesel" + 0.017*"silverado" + 0.012*"fullsize" + 0.012*"good
y" + 0.011*"running" + 0.010*"ticket" + 0.010*"soso" + 0.010*"punch" + 0.009*"tradein"')
(5, '0.067*"mile" + 0.041*"year" + 0.037*"great" + 0.034*"good" + 0.028*"problem" + 0.026
*"reliable" + 0.025*"buy" + 0.025*"tire" + 0.020*"drive" + 0.019*"change"')
(6, '0.012*"work" + 0.011*"drive" + 0.011*"like" + 0.010*"control" + 0.008*"wheel" + 0.007
*"road" + 0.007*"system" + 0.007*"tire" + 0.006*"time" + 0.006*"speed"')
(7, '0.016*"pearl" + 0.015*"nissian" + 0.014*"bug" + 0.012*"advice" + 0.011*"white" + 0.01
0*"serpentine" + 0.009*"paper" + 0.009*"job" + 0.008*"encourage" + 0.008*"combo"')
(8, '0.051*"love" + 0.043*"drive" + 0.034*"great" + 0.021*"trip" + 0.017*"like" + 0.015*"s
eat" + 0.014*"mileage" + 0.014*"comfortable" + 0.013*"room" + 0.012*"mile"')
(9, '0.106*"truck" + 0.023*"great" + 0.022*"nissan" + 0.021*"frontier" + 0.019*"good" + 0.
017*"power" + 0.017*"drive" + 0.017*"vehicle" + 0.015*"well" + 0.014*"look"')
(10, '0.071*"bose" + 0.068*"system" + 0.052*"sound" + 0.038*"crew" + 0.037*"stereo" + 0.02
3*"corolla" + 0.018*"commuter" + 0.014*"leather" + 0.014*"speaker" + 0.012*"fosgate"')
(11, '0.032*"drive" + 0.026*"nissan" + 0.019*"altima" + 0.014*"vehicle" + 0.014*"buy" + 0.
014*"purchase" + 0.013*"mile" + 0.010*"like" + 0.009*"love" + 0.009*"month"')
(12, '0.035*"drive" + 0.032*"great" + 0.032*"good" + 0.023*"look" + 0.023*"maxima" + 0.021
*"love" + 0.020*"like" + 0.015*"sport" + 0.014*"car" + 0.013*"performance"')
(13, '0.038*"seat" + 0.032*"interior" + 0.026*"great" + 0.023*"good" + 0.023*"versa" + 0.0
19*"comfortable" + 0.017*"nice" + 0.016*"ride" + 0.013*"power" + 0.012*"room"')

 Mallet Coherence Score:  0.492988568011905
Wall time: 5min 19s
```

In [32]:
```
def my_coherence_vals(dictionary, corpus, texts, limit, start, step):
    coherence_values = []
    model_list = []
    for num_topics in range(start, limit, step):
        model = LdaMallet(mallet_path, corpus=corpus_1, num_topics=num_topics, id2word=id2
        model_list.append(model)
        coherencemodel = CoherenceModel(model=model, texts=texts, dictionary=dictionary, c
        coherence_values.append(coherencemodel.get_coherence())

    return model_list, coherence_values
```

In [33]:
```
# To get the coherence values
model_list, coherence_values = my_coherence_vals(dictionary=id2word_1, corpus=corpus_1,
                                                 texts=clean_brand_review, start=2, limit=
```

In [34]:
```
# Show graph for the coherence value scores vs number of topics
limit=26; start=2; step=6;
topics = range(start, limit, step)
plt.plot(topics, coherence_values)
plt.title("Coherence value score with the number of topics")
plt.xlabel("Num Topics")
plt.ylabel("Coherence score")
plt.legend(("coherence_values"), loc='best')
plt.show()
```



In [35]:
```
# Print the coherence scores
for best, cv in zip(topics, coherence_values):
    print("Topic ", best, " has Coherence Value of", round(cv, 4))
```

```
Topic  2  has Coherence Value of 0.4358
Topic  8  has Coherence Value of 0.4635
Topic  14  has Coherence Value of 0.4794
Topic  20  has Coherence Value of 0.4701
```

In [36]:
```
# printing the best topics
optimal_model = model_list[1]
model_topics = optimal_model.show_topics(formatted=False)
pprint(optimal_model.print_topics(num_words=10))
```

```
[(0,
  '0.118*"nissan" + 0.057*"problem" + 0.043*"transmission" + 0.036*"issue" + '
  '0.030*"dealer" + 0.024*"replace" + 0.021*"warranty" + 0.019*"repair" + '
  '0.019*"dealership" + 0.018*"time"'),
 (1,
  '0.049*"drive" + 0.033*"month" + 0.028*"work" + 0.025*"mile" + 0.024*"time" '
  '+ 0.020*"find" + 0.017*"trip" + 0.017*"week" + 0.016*"long" + '
  '0.012*"purchase"'),
 (2,
  '0.084*"love" + 0.074*"drive" + 0.073*"great" + 0.032*"ride" + 0.023*"feel" '
  '+ 0.021*"comfortable" + 0.021*"road" + 0.020*"performance" + 0.019*"smooth" '
  '+ 0.019*"price"'),
 (3,
  '0.110*"truck" + 0.060*"great" + 0.043*"power" + 0.042*"good" + '
  '0.020*"frontier" + 0.018*"road" + 0.017*"titan" + 0.016*"toyota" + '
  '0.012*"size" + 0.012*"ride"'),
 (4,
  '0.058*"good" + 0.039*"drive" + 0.037*"mileage" + 0.029*"altima" + '
  '0.028*"highway" + 0.026*"car" + 0.025*"maxima" + 0.023*"fuel" + '
  '0.022*"speed" + 0.019*"sentra"'),
 (5,
  '0.053*"seat" + 0.020*"system" + 0.020*"interior" + 0.017*"nice" + '
  '0.016*"rogue" + 0.015*"control" + 0.015*"easy" + 0.014*"feature" + '
  '0.013*"small" + 0.012*"room"'),
 (6,
  '0.034*"tire" + 0.031*"brake" + 0.026*"replace" + 0.021*"noise" + '
  '0.019*"door" + 0.018*"rear" + 0.017*"turn" + 0.016*"engine" + 0.015*"light" '
  '+ 0.014*"thing"'),
 (7,
  '0.079*"vehicle" + 0.079*"year" + 0.076*"mile" + 0.071*"buy" + '
  '0.032*"purchase" + 0.027*"reliable" + 0.023*"pathfinder" + 0.019*"nissan" + '
  '0.019*"problem" + 0.018*"money"')]
```

Topic 1 -->> Problem with dealers and services Topic 2 -->> Topic 3 -->> Love driving the car Topic 4 -->> Car performance and quality Topic 5 -->> Speed and fuel consumption Topic 6 -->> how comfortable the car seat

Topic 7 -->> The problem with maintenance and warranty Topic 8 -->> Problems with mileage and cost

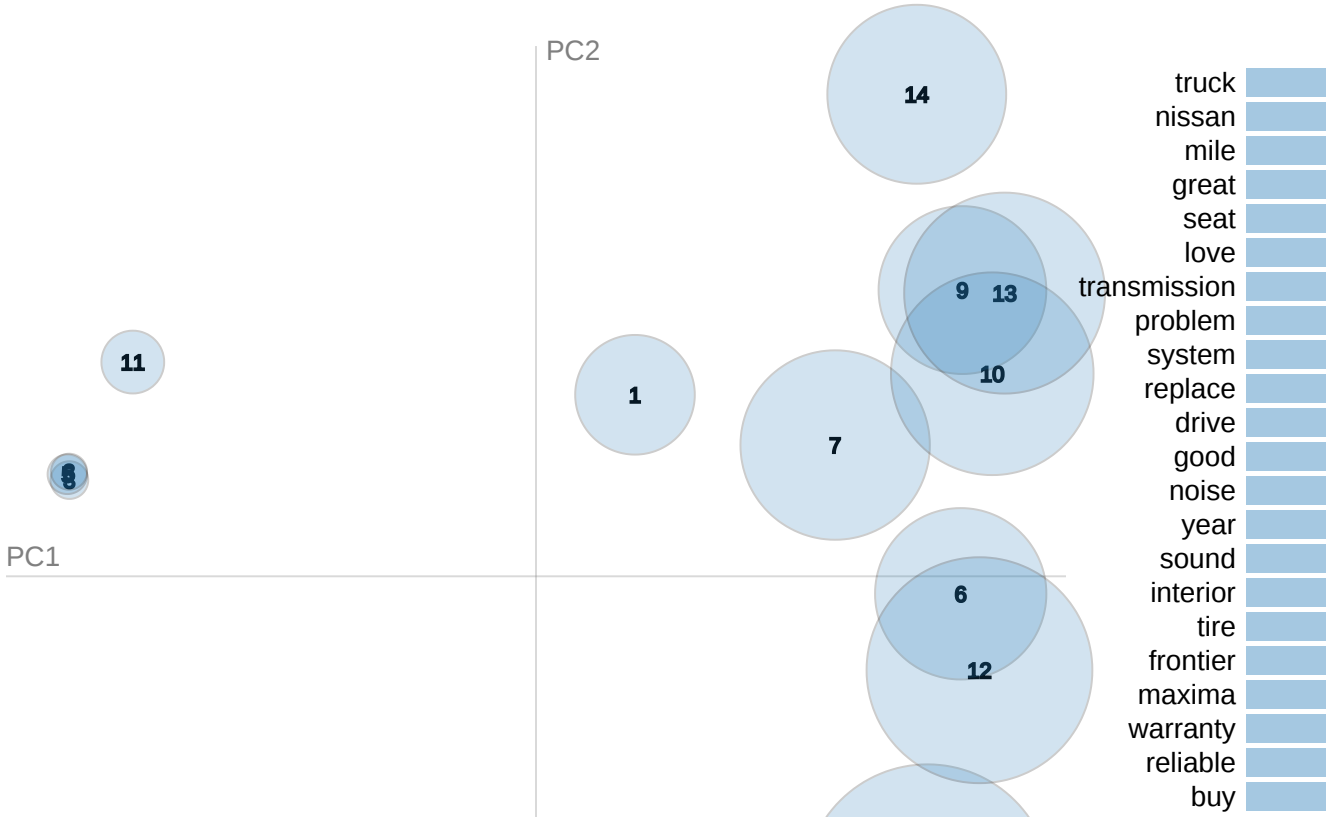
In [37]:
```
#Visualize the topic
pyLDAvis.enable_notebook()
LDAvis_prepared = pyLDAvis.gensim.prepare(ldamallet, corpus=corpus_1, dictionary=id2word_1
LDAvis_prepared
```

Out[37]:

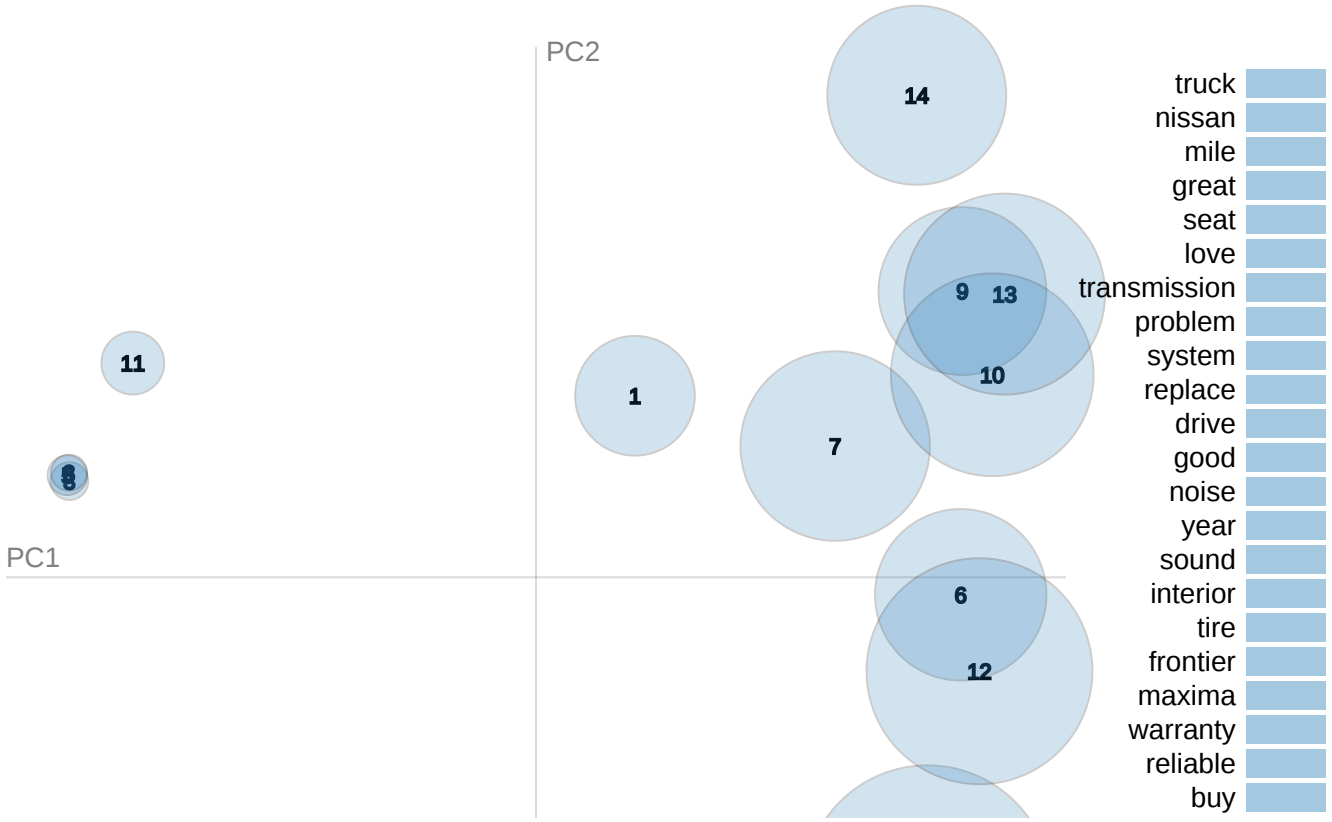
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to a
metric:(2) Slid
metric:(2)

Intertopic Distance Map (via multidimensional scaling)



Selected Topic: 0 Previous Topic Next Topic Clear Topic



In [38]:
```
# Save the visulaization to html
pyLDAvis.save_html(LDAvis_prepared, 'LdaModel_viz.html')
```

In [58]:
```
# Define the sentence topics
def sentence_topics(ldamodel=ldamodel, corpus=corpus_1, texts=clean_brand_review):
    # Init output
    topics_df = pd.DataFrame()

    # Looping through the documents to find the main topics
    for i, row in enumerate(ldamodel[corpus]):
        row = sorted(row[0], key=lambda x: (x[1]), reverse=True)

        # look for the Dominant topic, % contribution and Keywords
        for j, (topic_num, prop_topic) in enumerate(row):

            # Diplay the dominant topics
            if j == 0:
                dom = ldamodel.show_topic(topic_num)
                topic_keywords = ", ".join([word for word, prop in dom])
                topics_df = topics_df.append(pd.Series([int(topic_num), round(prop_topic,2
            else:
                break
    topics_df.columns = ['Dominant_Review_Topic', 'Perc_Contribution', 'Topic_Keywords']

    # Concatenate the text and the topics_df
    contents = pd.Series(texts)
    topics_df = pd.concat([topics_df, contents], axis=1)
    return(topics_df)


df_topic_sents_keywords = sentence_topics(ldamodel=ldamodel, corpus=corpus_1, texts=clean_

# Format
dominant_review_topic = df_topic_sents_keywords.reset_index()
dominant_review_topic.columns = ['Review_No', 'Dominant_Review_Topic', 'Percent_contr_per_
```

```
# Show
dominant_review_topic
```

Out[58]:

	Review_No	Dominant_Review_Topic	Percent_contr_per_topic	Review_Keywords	Original review
0	0	5.0	47.999999	drive, love, like, look, nissan, good, quest, great, vehicle, want	[outstanding, large, family, expect, arrival, child, toyota, sienna, minivan, go, small, need, thought, dive, huge, passenger, appeal, choice, long, time, pretty, ford, chevy, limit, feature, mercedes, nice, expensive, pleased, learn, nissan, start, sell, passenger, van, price, ford, chevy, van,...
1	1	0.0	87.000000	seat, good, like, rear, interior, nice, door, great, drive, turn	[suck, rear, blow, slow, want, seat]
2	2	2.0	89.999998	great, drive, love, good, look, altima, like, nissan, ride, comfortable	[love, small, astro, typeyou, need, navigate, certain, parking, lotsspace, drive, fast, food, drive, thrus, buy, extra, roomey]
3	3	5.0	60.000002	drive, love, like, look, nissan, good, quest, great, vehicle, want	[nissan, review, satisfied, nissan, business, delivery, personal, camping, road, trip, child, store, seat, warehouse, want, passenger, rear, conditioning, drive, florida, california, cross, country, trip, average, drive, rain, comfortable, stable, vehicle, nissan, titan, engine, mile, engine, te...
4	4	0.0	51.999998	seat, good, like, rear, interior, nice, door, great, drive, turn	[family, go, honda, odyssey, family, grow, need, room, kid, booster, seat, pro, smooth, drivecomfortable, driver, configuration, shift, stick, get, adjust, handle, road, great, warranty, people, lanecon, wide, turn, difficult, maneuver, shopping, parking, lot, carpool, lane, school, power, long,...
...	...	...	...	...	...

	Review_No	Dominant_Review_Topic	Percent_contr_per_topic	Review_Keywords	Original review
11710	11710	2.0	56.999999	great, drive, love, good, look, altima, like, nissan, ride, comfortable	[build, credit, nissan, versa, march, year, big, complaint, dealership, sell, review, major, dallas, dealership, review, rent, car, rental, drive, versa, love, drive, rental, city, average, mile, go, shortly, buy, buy, awesome, talk, handle, great, response, tight, steer, great, acceleration, gr...
11711	11711	5.0	69.999999	drive, love, like, look, nissan, good, quest, great, vehicle, want	[good, elderly, like, dealer, want, honor, certificate, bernardino, nissan, wernt, financing, eventually, pete, arrange]
11712	11712	10.0	47.999999	xterra, drive, love, good, sport, great, look, road, like, performance	[warrant, drive, outside, condenser, cover, apparently, common, damage, rock, roadprotecte, underneathbut, frontwhich, amazingly, open, unprotecte, clearly, design, flaw]
11713	11713	7.0	38.999999	mile, drive, mileage, trip, great, highway, average, vehicle, long, city	[reliable, transportation, sport, reliable, transportation, price, handle, winter, weather, condition, city, drive]
11714	11714	2.0	68.000001	great, drive, love, good, look, altima, like, nissan, ride, comfortable	[versa, go, nissan, look, kind, afford, expect, walk, suggest, learn, versa, month, comfortable, storage, space, roomy, especially, seat, engine, noisy, run, smoothly, problem, manuevere, traffic, parking, spot, understand, subcompact, vehicle, feel, little, tight, tall, person, drive, comfortab...

11715 rows × 5 columns

In [48]:

```
# The Dataframe
sent_topics_df = pd.DataFrame()

topics_out = df_topic_sents_keywords.groupby('Dominant_Review_Topic')

for i, j in topics_out:
    sent_topics_df = pd.concat([sent_topics_df,j.sort_values(['Perc_Contribution'], ascend


sent_topics_df.reset_index(drop=True, inplace=True)

# Format
sent_topics_df.columns = ['Topic_Num', "Percent_contr_per_topic", "Review_Keywords", "Orig
```

```
# Display the 8 topics
sent_topics_df
```

Out[48]:

	Topic_Num	Percent_contr_per_topic	Review_Keywords	Original review
0	0.0	99.000001	seat, good, like, rear, interior, nice, door, great, drive, turn	[subcompact, think, corolla, week, place, hertz, sale, center, approx, mile, approx, dealer, want, thousand, high, mileage, one, handle, like, dream, need, light, touch, accelerate, fine, city, highway, love, small, touch, wellplaced, bottle, holder, glove, compartment, huge, hold, woman, pocket...
1	1.0	98.000002	vehicle, pathfinder, good, rogue, well, ride, armada, feature, like, feel	[like, fly, firstclass, design, armada, set, apart, competition, move, away, boxy, sameness, american, automaker, past, year, armada, stand, refreshing, design, great, road, presence, comfort, driving, excellent, generous, interior, superb, quiet, highway, speed, excellent, acceleration, capabil...
2	2.0	98.000002	great, drive, love, good, look, altima, like, nissan, ride, comfortable	[good, special, buy, family, friend, year, mile, major, problem, feel, like, midrange, aspect, instance, engine, responsive, like, sport, coupe, get, decent, mileage, spectacular, like, look, design, interior, material, make, feel, little, cheap, overall, think, good, thing, great, look, feel, s...
3	3.0	99.000001	truck, great, frontier, good, nissan, titan, drive, power, like, look	[truck, look, oofficedocumentsettings, oallowpng, oofficedocumentsettingsxmlendifif, wworddocument, wviewnormalwview, wzoomwzoom, wtrackmove, wtrackformatte, wpunctuationkerne, wvalidateagainstschemas, wsaveifxmlinvalidfalsewsaveifxmlinvalid, wignoremixedcontentfalsewignoremixedcontent, walwayss...
4	4.0	98.000002	nissan, year, need, look, clutch, purchase, like, content, go, truck	[bring, door, styling, promise, husband, replace, alloy, wheel, subtle, change, appearancce, wheel, well, means, look, like, flamingwrite, ticket, officer, sleeper, stealthy, nissan, listen, hellooo, bring, new, drivetrainengine, sentra, today, limited, edition, prostitute, flame, model, indicat...
5	5.0	99.000001	drive, love, like, look, nissan, good, quest, great, vehicle, want	[melt, dollop, cream, cube, own, white, nissan, cube, week, summertime, miata, prefer, ride, cube, love, complaint, somewhat, questionable, acceleration, true, electronic, seemingly, make, decision, say, get, issue, highway, miata, like, tightly, laced, sprinting, shoe, cube, like, pair, flipflo...
6	6.0	97.000003	wheel, rear, turn, work, good, radius, roof, rack, driver, truck	[week, againjust, crush, stonewas, able, pull, effortlesslythe, truck, sink, hardly, bucket, loader, fill, trailerour, drop, tail, look, like, rear, wheel, rubi, need, large, drop, hitch, trailer, angle, come, drop, rigstay, tune]
7	7.0	98.000002	mile, drive, mileage, trip, great, highway, average, vehicle, long, city	[fast, charger, need, mile, month, great, purchase, tired, wait, smart, repeat, delayswe, base, model, fast, charge, optionsl, model, come, onboard, charger, charge, hourthe, charger, standard, give, half, charge, rate, meaning, force, wait, long, charge, townthis, huge, dealthe, fast, charge, o...
8	8.0	95.999998	head, engine, gasket, like, look, good, thing, go, build, car	[scan, button, radio, push, button, wait, juke, out, scan, button, , fire, want, money, wait, late, , moron, check, scan, button]
9	9.0	98.000002	juke, love, problem, vehicle, year, sentra, drive, buy, nissan, time	[meet, exceed, expectation, receive, risk, purchase, base, hype, rumour, review, able, testdrive, exceed, expectation, incredibly, drive, course, power, expect, expect, awesome, double, clutch, automanual, transmission, shift, quickly, precisely, problem, passenger, door, align, close, properly,...

	Topic_Num	Percent_contr_per_topic	Review_Keywords	Original review
10	10.0	98.000002	xterra, drive, love, good, sport, great, look, road, like, performance	[greatgreat, drive, decide, spec, think, worth, money, plus, dealership, willing, come, price, spec, seat, great, love, confidence, speed, turn, butt, stay, plant, seat, slide, previously, drive, pontiac, grand, prix, acceleration, stunning, gear, love, standard, feature, great, power, handling]
11	11.0	99.000001	nissan, problem, transmission, mile, dealer, time, replace, go, warranty, issue	[purge, valve, connector, tranamission, go, patterson, nissan, longview, check, engine, light, find, broken, wire, purge, valve, connnector, day, go, dealer, contact, repair, dealer, want, broken, wire, connector, say, wire, harness, replace, real, say, comon, problem, nissan, mention, technical...
12	12.0	99.000001	mile, year, replace, tire, drive, buy, problem, good, great, brake	[good, own, buy, mile, problem, engine, transmission, mobil, synthetic, weight, close, tolerance, machining, japanese, engine, transmission, flush, power, drive, gentle, high, mileage, thing, like, pull, right, time, little, think, acceleration, wheel, drive, tire, laser, alignment, slight, pull...
13	13.0	98.000002	paint, nissan, vehicle, xterra, quality, rust, year, bumper, scratch, chip	[beware, paint, redline, own, yeari, willing, deal, minor, issue, year, see, nissan, north, america, supportive, fix, issue, come, screache, halt, paint, come, asflake, leave, white, spot, entire, carthis, mile, garage, pampered, look, terriblenissan, north, america, state, rock, damage, rectify...

In [41]:
```
#saving the review of each brand in a DataFrame
Altima_df = ([sent for sent in brand.loc[brand['model'] == 'Altima', 'review']])
Sentra_df = ([sent for sent in brand.loc[brand['model'] == 'Sentra', 'review']])
Frontier_df = ([sent for sent in brand.loc[brand['model'] == 'Frontier', 'review']])
Maxima_df = ([sent for sent in brand.loc[brand['model'] == 'Maxima', 'review']])
Pathfinder_df = ([sent for sent in brand.loc[brand['model'] == 'Pathfinder', 'review']])
Titan_df = ([sent for sent in brand.loc[brand['model'] == 'Titan', 'review']])
Xterra_df = ([sent for sent in brand.loc[brand['model'] == 'Xterra', 'review']])
Rogue_df = ([sent for sent in brand.loc[brand['model'] == 'Rogue', 'review']])
Versa_df = ([sent for sent in brand.loc[brand['model'] == 'Versa', 'review']])

Quest_df = ([sent for sent in brand.loc[brand['model'] == 'Quest', 'review']])
Armada_df = ([sent for sent in brand.loc[brand['model'] == 'Armada', 'review']])
Juke_df = ([sent for sent in brand.loc[brand['model'] == 'Juke', 'review']])

Leaf_df = ([sent for sent in brand.loc[brand['model'] == 'Leaf', 'review']])

Cube_df = ([sent for sent in brand.loc[brand['model'] == 'Cube', 'review']])

Truck_df = ([sent for sent in brand.loc[brand['model'] == 'Truck', 'review']])

NV200_df = ([sent for sent in brand.loc[brand['model'] == 'NV200', 'review']])
Murano_df = ([sent for sent in brand.loc[brand['model'] == 'Murano', 'review']])

NV_df = ([sent for sent in brand.loc[brand['model'] == 'NV', 'review']])
Kicks_df = ([sent for sent in brand.loc[brand['model'] == 'Kicks', 'review']])
Z350_df = ([sent for sent in brand.loc[brand['model'] == 'Z350', 'review']])
Z370_df = ([sent for sent in brand.loc[brand['model'] == 'Z370', 'review']])
SX200_df = ([sent for sent in brand.loc[brand['model'] == 'SX200', 'review']])
SX240_df = ([sent for sent in brand.loc[brand['model'] == 'SX240', 'review']])
GTR_df = ([sent for sent in brand.loc[brand['model'] == 'GTR', 'review']])
```

In [49]:
```
%%time
# Defining a function to get the topics and visualize them
def each_brand(text):

    # Create Dictionary
```

```
    id2word_2 = corpora.Dictionary(text)

    # Create Corpus: Term Document Frequency
    corpus_2 = [id2word_2.doc2bow(review) for review in text]

    # Here I decided to reduce the number of topics to only six for each brand
    model = LdaMulticore(corpus=corpus_2, num_topics = 8, id2word=id2word_2,chunksize=2000

    LDAvis_prepared = pyLDAvis.gensim.prepare(model, corpus=corpus_2, dictionary=id2word_2

    return LDAvis_prepared
```

```
Wall time: 0 ns
```

In [57]:
```
brand.head()
```

Out[57]:

	Rating	review	year	car_name	model	date	review_year	month	day
0	4	[outstanding, large, family, expect, arrival, child, toyota, sienna, minivan, go, small, need, thought, dive, huge, passenger, appeal, choice, long, time, pretty, ford, chevy, limit, feature, mercedes, nice, expensive, pleased, learn, nissan, start, sell, passenger, van, price, ford, chevy, van,...	2013	Nissan	NV	2013-03-07	2013	3	7
1	3	[suck, rear, blow, slow, want, seat]	2015	Nissan	NV	2018-07-06	2018	7	6
2	5	[love, small, astro, typeyou, need, navigate, certain, parking, lotsspace, drive, fast, food, drive, thrus, buy, extra, roomey]	2015	Nissan	NV	2018-03-26	2018	3	26
3	5	[nissan, review, satisfied, nissan, business, delivery, personal, camping, road, trip, child, store, seat, warehouse, want, passenger, rear, conditioning, drive, florida, california, cross, country, trip, average, drive, rain, comfortable, stable, vehicle, nissan, titan, engine, mile, engine, te...	2015	Nissan	NV	2016-05-14	2016	5	14
4	3	[family, go, honda, odyssey, family, grow, need, room, kid, booster, seat, pro, smooth, drivecomfortable, driver, configuration, shift, stick, get, adjust, handle, road, great, warranty, people, lanecon, wide, turn, difficult, maneuver, shopping, parking, lot, carpool, lane, school, power, long,...	2015	Nissan	NV	2015-10-21	2015	10	21

In [52]:
```
GTR_lda = each_brand(GTR_df)
```

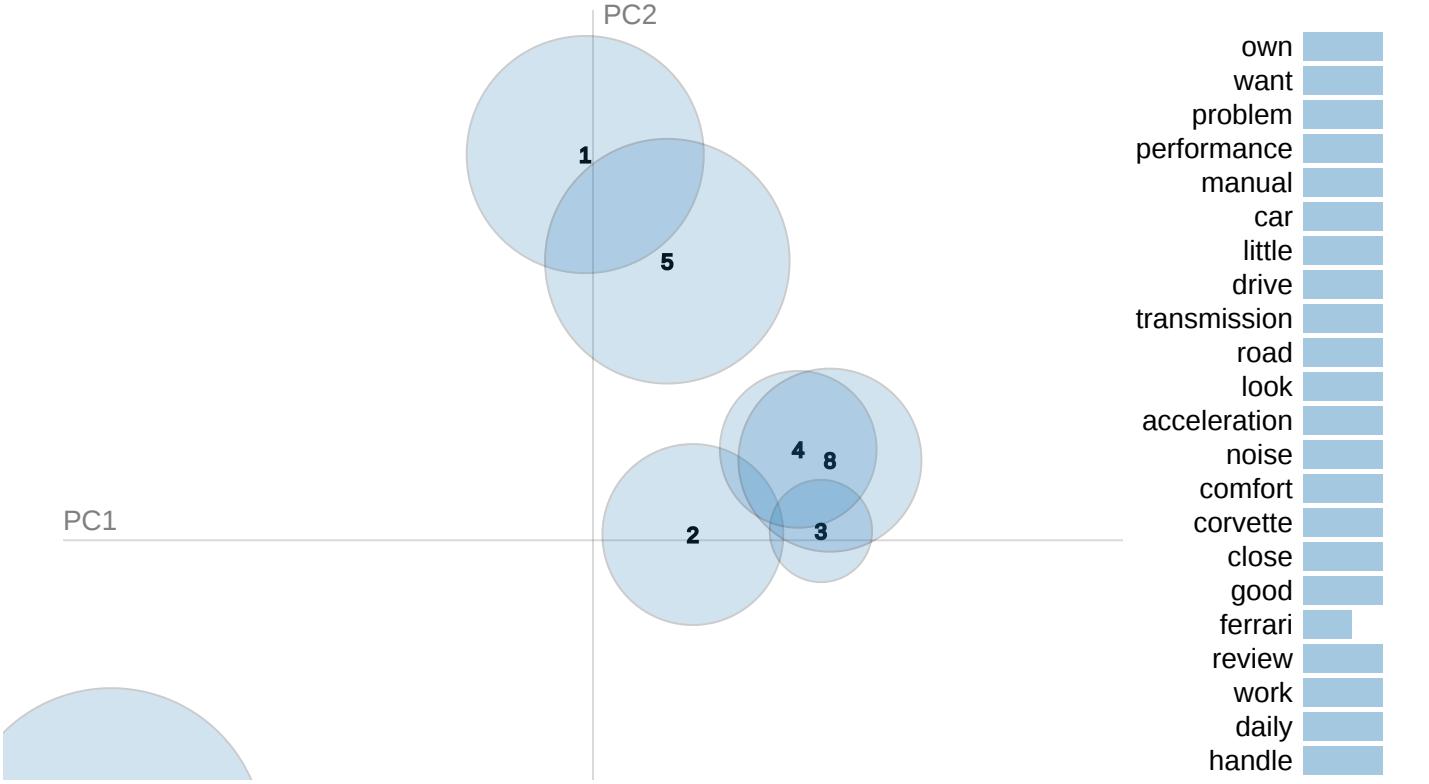
In [53]:
```
GTR_lda
```

Out[53]:

Selected Topic: [ ] Previous Topic Next Topic Clear Topic

Slide to a
metric:(2)

Slid
metric:(2)



Topic 1 -->> Problem with dealers and services Topic 2 -->> Topic 3 -->> Love driving the car Topic 4 -->> Car performance and quality Topic 5 -->>Speed and fuel consumption Topic 6 -->> how comfortable the car seat Topic 7 -->> The problem with maintenance and warranty Topic 8 -->> Problems with mmileage and cost

In [54]:
```
NV200_lda = each_brand(NV200_df)
```

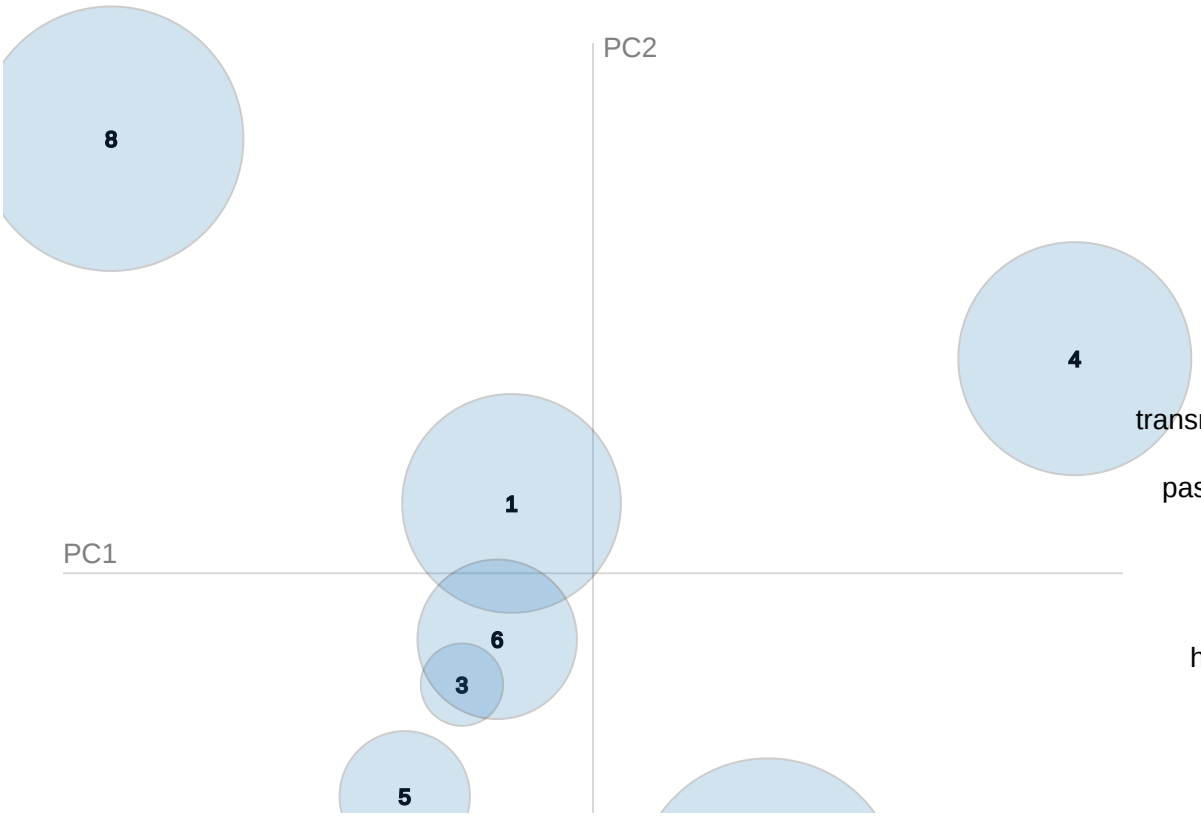
In [56]:
```
NV200_lda
```

Out[56]:

Selected Topic: [ ] Previous Topic Next Topic Clear Topic

Slide to a
metric:Sli(2)
metric:(2)

Intertopic Distance Map (via multidimensional scaling)



Selected Topic: 0 Previous Topic Next Topic Clear Topic

Intertopic Distance Map (via multidimensional scaling)