

Natural Language Processing final Project

Topic Analysis of Review Data - VOC

Reviews Analysis of the Newly lunched Lenova Mobile in India on Amazon.in

In [57]:

```
import pandas as pd
import numpy as np
import nltk
import seaborn as sns
import matplotlib.pyplot as plt
```

In [58]:

```
df=pd.read_csv('K8_Reviews_v0.2.csv')
df.head()
```

Out[58]:

	sentiment	review
0	1	Good but need updates and improvements
1	0	Worst mobile i have bought ever, Battery is dr...
2	1	when I will get my 10% cash back.... its alrea...
3	1	Good
4	0	The worst phone everThey have changed the last...

In [59]:

```
df.columns
```

Out[59]:

```
Index(['sentiment', 'review'], dtype='object')
```

In [60]:

```
df.sentiment.value_counts()
```

Out[60]:

```
0    7712
1    6963
Name: sentiment, dtype: int64
```

Columns: Sentiment: The sentiment against the review (4,5 star reviews are positive, 1,2 are negative)

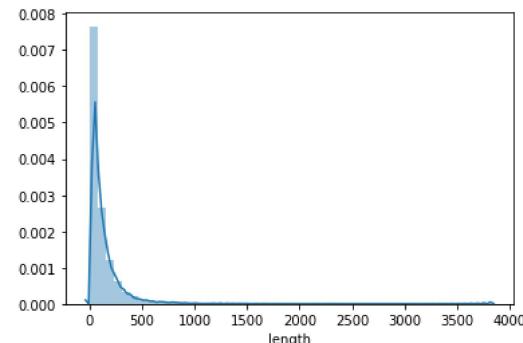
Reviews: The main text of the review

In [61]:

```
# getting the Length of the reviews
df['length']=df['review'].apply(len)
sns.distplot(df.length)
```

Out[61]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1816f7951c8>
```

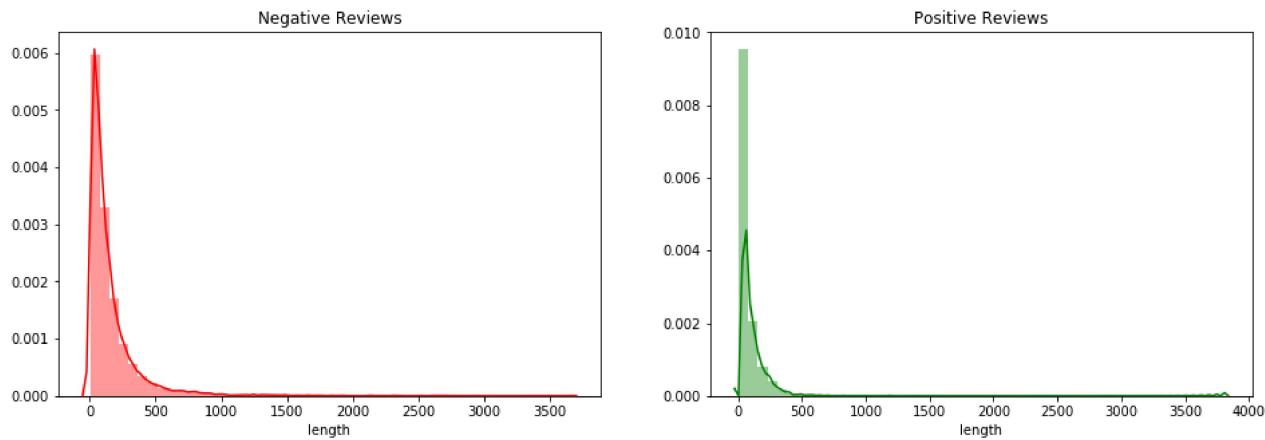


The average length of the reviews are well within 500 words

Visualization showing the distributions of the sentiments of each review

In [62]:

```
fig,(ax1,ax2)=plt.subplots(ncols=2,nrows=1,figsize=(16,5))
sns.distplot(df[df.sentiment ==0].length, color='red',ax=ax1)
ax1.set_title('Negative Reviews')
sns.distplot(df[df.sentiment ==1].length, color='green',ax=ax2)
ax2.set_title('Positive Reviews')
plt.show()
```



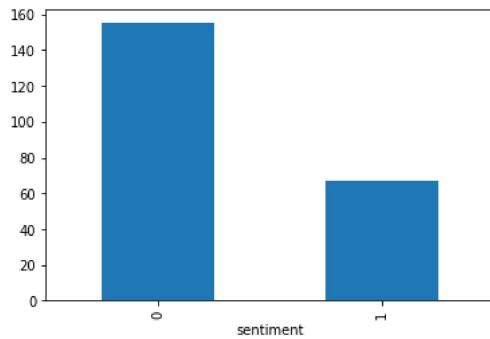
Average lengths of the reviews when they are Positive and Bad

In [63]:

```
grouped_df=df.groupby('sentiment')
grouped_df.length.mean().plot(kind='bar')
```

Out[63]:

<matplotlib.axes._subplots.AxesSubplot at 0x1816424f3c8>



It shows that customers write longer and detailed reviews when they are not much satisfied with the product

NLP analysis

Tokenizing the reviews

first we need to convert the sentiments in array format into the list so we can tokenize them.

In [64]:

```
reviews=list(df.review.values)
#converting into Lower
reviews=[review.lower().split('.') for review in reviews]
```

In [65]:

```
print(reviews[0:10])
```

```
[['good but need updates and improvements'], ['worst mobile i have bought ever, battery is draining like hell, backup is only 6 to 7 hour mobile idle its getting discharged', 'this is biggest lie from amazon & lenovo which is not at all expected, they are making full by saying charger is fake, it takes at least 4 to 5 hours to be fully charged', "don't know how lenovo will survive by making full of us", 'please let like me', ''], ['when i will get my 10% cash back', ' ', ' ', ' ', ' its already 15 january', ' ', ''], ['good'], ['the worst phone ever the problem is still same and the amazon is not returning the phone ', 'highly disappointing of amazon'], ["only i'm telling don't buy i'm a camera waste of money"], ['phone is awesome', ' but while charging, it heats up allot', ' ', 'really a genuine reason to hate lenovo k8 note wn'], ["it's over hitting problems", ' ', ' ', 'and phone hanging problems lenovo k 8 note', ' ', ' ', "so where is service station in ahmedabad can change the phone by lenovo"], ['a lot of glitches dont buy this thing better go for some other options']]
```

Tokenizing using word tokenizer of nltk package

In [66]:

```
token=nltk.WhitespaceTokenizer()
review_tokens = [token.tokenize(review[0]) for review in reviews]
print(review_tokens[0:5])
```

```
[['good', 'but', 'need', 'updates', 'and', 'improvements'], ['worst', 'mobile', 'i', 'have', 'bought', 'ever', 'battery', 'is', 'draining', 'only', '6', 'to', '7', 'hours', 'with', 'internet', 'uses', 'even', 'if', 'i', 'put', 'mobile', 'idle', 'its', 'getting', 'discharged'] 10%, 'cash', 'back'], ['good'], ['the', 'worst', 'phone', 'everthey', 'have', 'changed', 'the', 'last', 'phone', 'but', 'the', 'problem', 'amazon', 'is', 'not', 'returning', 'the', 'phone']]
```

Perform parts-of-speech tagging on each sentence using the NLTK POS tagger.

In [67]:

```
review_pos=[nltk.pos_tag(review) for review in review_tokens]
#review_pos=[nltk.pos_tag(review,tagset='universal') for review in review_tokens]
review_pos[0]
```

Out[67]:

```
[('good', 'JJ'), ('but', 'CC'), ('need', 'VBP'), ('updates', 'NNS'), ('and', 'CC'), ('improvements', 'NNS')]
```

In [236]:

```
#print(nltk.help.upenn_tagset())
```

Extracting the nouns form the POS tagged reviews as per problem statement

In [69]:

```
review_nouns=[]
for n in review_pos:
    review_nouns.append([postag for postag in n if postag[1].startswith('NN')])
#review_nouns.append([postag for postag in n if postag[1]=='NOUN'])
for n in range (10):
    print('Nouns in {} review --> {} \n'.format(n+1,review_nouns[n]))
```

Nouns in 1 review --> [('updates', 'NNS'), ('improvements', 'NNS')]

Nouns in 2 review --> [('mobile', 'NN'), ('i', 'NN'), ('battery', 'NN'), ('hell,', 'NN'), ('backup', 'NN'), ('hours', 'NNS'), ('internet', 'NN')]

Nouns in 3 review --> [('i', 'NN'), ('cash', 'NN')]

Nouns in 4 review --> []

Nouns in 5 review --> [('phone', 'NN'), ('everthey', 'NN'), ('phone', 'NN'), ('problem', 'NN'), ('amazon', 'NN'), ('phone', 'NN')]

Nouns in 6 review --> [("i'm", 'NN'), ("buyi'm", 'NN'), ('camerawaste', 'NN'), ('money', 'NN')]

Nouns in 7 review --> [('phone', 'NN')]

Nouns in 8 review --> [('battery', 'NN'), ('level', 'NN')]

Nouns in 9 review --> [("it's", 'NN'), ('hitting', 'NN'), ('problems', 'NNS')]

Nouns in 10 review --> [('lot', 'NN'), ('glitches', 'NNS'), ('thing', 'NN'), ('options', 'NNS')]

Lemmatize the words extarcted

In [70]:

```
from nltk.stem import WordNetLemmatizer
lemmatizer=WordNetLemmatizer()
review_lem=[]
```

In [71]:

```
for review in review_nouns:
    review_lem.append([lemmatizer.lemmatize(postag[0], 'n') for postag in review])
print(review_lem[0:10])
```

[['update', 'improvement'], ['mobile', 'i', 'battery', 'hell,', 'backup', 'hour', 'internet', 'uses,', 'idle'], ['i', 'cash'], [], ['phone', 'amazon', 'phone'], ["i'm", "buyi'm", 'camerawaste', 'money'], ['phone'], ['battery', 'level'], ["it's", 'hitting', 'problem'], ['lot', 'glitch', 'thing']]

Removing the Stopwords and punctuation

In [72]:

```
from nltk.corpus import stopwords
from string import punctuation

clean_review=[]
stopwords=list(stopwords.words('english'))
punctuation=list(punctuation)
```

In [73]:

```
for words in review_lem:
    clean_review.append([word for word in words if word not in (stopwords or punctuation)])
```

In [74]:

```
print(clean_review[0:10])
```

[['update', 'improvement'], ['mobile', 'battery', 'hell,', 'backup', 'hour', 'internet', 'uses,', 'idle'], ['cash'], [], ['phone', 'evert'], ["i'm", "buyi'm", 'camerawaste', 'money'], ['phone'], ['battery', 'level'], ['hitting', 'problem'], ['lot', 'glitch', 'thing']]

Preparing topic model using LDA on the cleaned reviews

In [75]:

```
# Gensim
import gensim
import gensim.corpora as corpora
from gensim.utils import simple_preprocess
from gensim.models import CoherenceModel

# Plotting tools
import pyLDAvis
import pyLDAvis.gensim
```

In [76]:

```
# Create Dictionary
id2word = corpora.Dictionary(clean_review)

# Filter out words that occur less than 10 documents, or more than 50% of the documents.
id2word.filter_extremes(no_below=10, no_above=0.5)

print(id2word)

# Create Corpus
texts= clean_review
print('Corpus \t',clean_review[0])
```

Dictionary(406 unique tokens: ['improvement', 'update', 'backup', 'battery', 'hour']...)
Corpus ['update', 'improvement']

WE have 406 words left after cleaning out most of the unrelated words for our analysis

And created an corpora in name of TEXT

Creating the Bag of Words(bos)

In [77]:

```
print(id2word.token2id) #tokens with id
```

```
{"improvement": 0, "update": 1, "backup": 2, "battery": 3, "hour": 4, "internet": 5, "mobile": 6, "cash": 7, "amazon": 8, "phone": 9, "product": 10, "level": 13, "lot": 14, "option": 15, "thing": 16, "wrost": 17, "charger": 18, "damage": 19, "month": 20, "life": 21, "purchase": 22, "sc5, "card": 26, "product": 27, "range": 28, "specification": 29, "smartphone": 30, "speaker": 31, "camera": 32, "speed": 33, "cast": 34, "i7, "price": 38, "function": 39, "one": 40, "performance": 41, "r": 42, "call": 43, "disappointment": 44, "signal": 45, "quality": 46, "rain": 50, "k8": 51, "network": 52, "memory": 53, "superb": 54, "value": 55, "heating": 56, "apps": 57, "mobile": 58, "ok": 59, "can't": 60, "device": 64, "usage": 65, "ka": 66, "mark": 67, "weight": 68, "point": 69, "issue": 70, "system": 71, "delivery": 72, "week": 73, "note": 77, "phone": 78, "product": 79, "centre": 80, "handset": 81, "service": 82, "u": 83, "feature": 84, "refund": 85, "anyone": 86, "year": 90, "gaming": 91, "music": 92, "segment": 93, "drawback": 94, "process": 95, "turbo": 96, "experience": 97, "good": 98, "expectation": 102, "phn": 103, "display": 104, "model": 105, "type": 106, "support": 107, "need": 108, "hai": 109, "hang": 110, "hota": 111, "cougust": 114, "drain": 115, "mah": 116, "smooth": 117, "auto": 118, "configuration": 119, "touch": 120, "thanks": 121, "audio": 122, "battle": 125, "term": 126, "connection": 127, "etc": 128, "lack": 129, "pc": 130, "ye": 131, "mode": 132, "volta": 133, "a1": 134, "compare": 135, "138, "processor": 139, "selfie": 140, "facility": 141, "record": 142, "key": 143, "recorder": 144, "super": 145, "description": 146, "bak": 49, "thank": 150, "use": 151, "bit": 152, "bug": 153, "image": 154, "anything": 155, "complaint": 156, "future": 157, "reception": 158, "k": 161, "think": 162, "voice": 163, "sound": 164, "please": 165, "resolution": 166, "invoice": 167, "choice": 168, "center": 169, "k5": 173, "night": 174, "hell": 175, "bill": 176, "budget": 177, "series": 178, "worth": 179, "mob": 180, "buy": 181, "version": 182, "lo": 85, "box": 186, "phone": 187, "headphone": 188, "volume": 189, "build": 190, "default": 191, "flash": 192, "light": 193, "atmos": 194, "n": 197, "pro": 198, "fast": 199, "spec": 200, "clearity": 201, "gallery": 202, "everything": 203, "fone": 204, "waste": 205, "contact": 208, "backup": 209, "date": 210, "killer": 211, "till": 212, "video": 213, "jio": 214, "i": 215, "pls": 216, "sir": 217, "nahi": 218, "gud": 222, "people": 223, "phon": 224, "head": 225, "package": 226, "awesome": 227, "price": 228, "experience": 229, "company": 221, "rage": 233, "drop": 234, "minute": 235, "feel": 236, "front": 237, "bhi": 238, "koi": 239, "mat": 240, "core": 241, "deca": 242, "ram": 2245, "awsome": 246, "perfect": 247, "scratch": 248, "response": 249, "condition": 250, "redmi": 251, "number": 252, "request": 253, "dont": 257, "mast": 258, "moto": 259, "deal": 260, "cam": 261, "n": 262, "app": 263, "restart": 264, "customer": 265, "color": 266, "medium": 270, "beken": 271, "pack": 272, "fingerprint": 273, "camera": 274, "sensor": 275, "heat": 276, "get": 277, "button": 278, "g": 1, "piece": 282, "replacement": 283, "pic": 284, "bt": 285, "nothing": 286, "osm": 287, "market": 288, "change": 289, "jack": 290, "amount": 293, "storage": 294, "user": 295, "ghatiya": 296, "hi": 297, "design": 298, "set": 299, "strength": 300, "working": 301, "recording": 302, "t": 305, "k4": 306, "player": 307, "brand": 308, "today": 309, "mera": 310, "wow": 311, "policy": 312, "gb": 313, "exchange": 314, "vibe": 318, "finger": 318, "print": 319, "lock": 320, "android": 321, "doubt": 322, "suck": 323, "k": 324, "multi": 325, "photo": 326, "hanging": 327, "ver": 330, "bluetooth": 331, "drainage": 332, "hand": 333, "part": 334, "offer": 335, "capacity": 336, "rating": 337, "category": 338, "h": 41, "rest": 342, "care": 343, "money": 344, "cell": 345, "others": 346, "world": 347, "mirror": 348, "ho": 349, "issue": 350, "ki": 351, "aper": 354, "help": 355, "k6": 356, "size": 357, "show": 358, "shot": 359, "seller": 360, "team": 361, "word": 362, "discharge": 363, "ore": 366, "picture": 367, "draining": 368, "play": 369, "love": 370, "bahut": 371, "ke": 372, "raha": 373, "nhi": 374, "brightness": 375, "cha": 8, "ph": 379, "heating": 380, "buying": 381, "order": 382, "guy": 383, "k8note": 384, "performance": 385, "result": 386, "faulty": 387, "390, "quality": 391, "batter": 392, "interface": 393, "wise": 394, "bokeh": 395, "cool": 396, "space": 397, "processing": 398, "mp": 395, "et": 402, "ui": 403, "batter": 404, "camara": 405}
```

In [78]:

```
#vector form of te above
review_bos=[id2word.doc2bow(text) for text in texts]# creating an vector representation of the above i,e; craeting bos
print(review_bos)
```



```
[1], [(9, 1), (16, 1), (23, 1), (92, 1), (177, 1), (266, 1), (275, 1)], [(216, 1), (329, 1)], [(6, 1)], [(52, 1)], [(9, 1)], [(96, 1), (6, 1), (276, 1)], [(9, 1)], [], [(27, 1)], [(9, 1)], [], [(115, 1), (193, 1), (199, 1), (365, 1), (389, 1)], [], [], [(9, 1), (27, 1), (1, 1)], [(3, 1), (10, 1)], [(9, 1), (32, 1), (70, 1), (100, 1), (207, 1), (322, 1)], [(27, 1)], [(27, 1)], [(29, 1), (38, 1), (70, 1), (84, 9, 2), (32, 1), (42, 1), (50, 1), (51, 1), (53, 1), (286, 1), (313, 1)], [(10, 1)], [(6, 1)], [], [(12, 1), (27, 1), (55, 1)], [(9, 1)], [(22, 1)], [(9, 1)], [], [(3, 1), (4, 1), (18, 1), (96, 1), (100, 1), (133, 1)], [], [(228, 1)], [(381, 1)], [(9, 1), (54, 1)], [(23 (9, 1), (32, 2), (33, 1), (108, 1), (185, 1), (237, 1), (300, 1)], [(27, 1)], [(9, 1), (68, 1)], [(1, 1), (9, 1), (32, 1), (37, 2), (70, 9, 1)], [(32, 1), (46, 1)], [(32, 1)], [(173, 1)], [(6, 1)], [(100, 2), (273, 1), (275, 1)], [(6, 1), (32, 1), (149, 1)], [], [], [(27, 41, 1)], [(103, 1)], [(9, 1), (20, 1)], [(20, 1), (23, 1), (173, 1), (248, 1)], [(77, 1), (165, 1)], [], [(32, 1), (211, 1), (268, 1)], 1)], [], [(8, 1), (33, 1), (41, 1), (51, 1), (101, 1), (297, 1), (383, 1)], [(3, 1), (31, 1), (32, 1), (156, 1)], [], [(6, 1), (65, 1)], 1)], [(10, 2), (70, 1)], [], [], [(168, 1)], [], [(32, 1), (46, 1), (173, 1)], [(57, 1)], [(7, 1), (172, 1), (374, 1)], [(27, 1)], [(6, 1), (51, 1)], [(2, 1), (3, 1), (27, 1), (133, 1)], [(9, 1), (20, 1), (70, 1), (112, 1), (290, 1)], [(370, 1)], [(32, 1), (63, 1), (2, 2)], [(27, 1)], [(56, 1), (63, 1), (184, 1)], [(12, 1), (205, 1)], [(1, 1), (243, 1)], [(32, 1), (56, 1), (70, 1)], [(306, 1)], [(9, 1), 1)], [(9, 1), (37, 1), (46, 1), (308, 1)], [], [(9, 1)], [(27, 1)], [(30, 1), (84, 1), (177, 1)], [], [(12, 1), (205, 1)], [], [(3, 1), 1)], [(32, 1)], [], [(6, 1)], [(9, 1), (161, 1), (207, 1), (321, 1)], [(9, 1)], [(27, 1), (205, 1)], [(9, 1)], [], [(9, 1), 3, 1), (365, 1), (400, 1)], [], [(10, 1), (299, 1)], [], [(32, 1), (46, 1), (193, 1)], [(6, 1), (157, 1)], [(27, 1)], [(58, 1), (84, 1)], [(9, 1)], [(27, 1)], [], [(9, 1), (25, 1)], [(9, 1), (38, 1)], [(27, 1)], [(27, 1)], [(31, 1)], [(38, 1), (84, 1), (103, 1)], [(20, 1)], 1), (203, 1), (332, 1)], [(12, 1), (54, 1), (55, 1), (78, 1)], [], [(51, 1), (64, 1), (77, 1)], [(38, 1), (78, 1), (92, 1), (188, 1)], [1), (27, 1), (82, 1)], [(6, 1)], [(77, 2), (84, 1)], [(21, 1)], [(27, 1), (29, 1)], [(9, 1), (84, 1)], [(9, 1)], [(96, 1)], [(383, 1)], [], [(9, 1), (32, 1), (394, 1)], [(9, 1), (26, 1), (32, 1), (84, 1), (87, 1), (92, 1), (123, 1), (128, 1)], [(51, 1), (77, 1), 3), (21, 1)], [(3, 1), (342, 1)], [(9, 1), (45, 1), (87, 1)], [(9, 1)], [(9, 1), (226, 1)], [(9, 1)], [(27, 1), (152, 1)], [(3, 1), (9, 7, 1)], [(6, 1), (32, 1), (84, 1)], [(9, 1)], [(9, 2), (41, 1)], [(6, 1), (49, 1), (50, 1)], [(404, 1)], [], [(9, 1), (102, 1), (261, 1), 77, 1), (309, 1)], [(9, 1)], [(37, 1), (51, 1), (77, 1), (85, 3), (109, 2), (324, 3), (349, 2), (378, 1)], [(3, 1), (28, 1), (27, 1), (108, 89, 1), (337, 1)], [(32, 1), (164, 1)], [(9, 1), (141, 1), (206, 1)], [(9, 2), (14, 1)], [(27, 1)], [], [(1, 1), (51, 1), (71, 1), (210, 74, 1), (309, 1)], [(1, 1), (32, 1), (237, 1), (389, 1)], [(9, 1), (27, 1)], [(2, 1), (3, 1), (25, 1), (139, 1)], [(6, 1)], [(203, 1), (209, 1)], [(9, 1)], [(6, 1), (9, 1), (20, 1), (36, 1), (77, 1), (181, 1)], [(77, 1), (211, 1)], [], [(77, 1)], [], [(10, 1), (205, 1)], [(6, 1), (27, 1), (107, 1), (113, 1), (216, 1), (265, 1)], [(12, 1)], [(10, 1), (52, 1), (127, 1)], [(9, 1)], [], [(3, 1), (1, 1), (77, 1)], [], [(9, 1), (32, 1), (398, 1)], [(27, 1)]]
```

In [79]:

```
print('Number of unique tokens: %d' % len(id2word))
print('Number of documents: %d' % len(review_bos))
```

Number of unique tokens: 406
Number of documents: 14675

In [80]:

```
# Build LDA model with 12 Topics
maxv=0
maxrs=0
from gensim.models import LdaModel

for rs in range(1,30):

    lda_model = LdaModel(corpus=review_bos,id2word=id2word,
                          num_topics=12,random_state=rs,passes=20,iterations=15)
    coherence_model_lda = CoherenceModel(model=lda_model, texts=clean_review, dictionary=id2word, coherence='c_v')
    coherence_lda = coherence_model_lda.get_coherence()

    print ('Coherence Score:{} RS: {}'.format(coherence_lda,rs))
    if coherence_lda > maxv:
        maxv = coherence_lda
        maxrs = rs

#print ("The ideal coherence score {:.4f} corresponds to random state {}".format(maxv,maxrs))

Coherence Score:0.5083919657696322 RS: 1
Coherence Score:0.5254187672082519 RS: 2
Coherence Score:0.476476041334444 RS: 3
Coherence Score:0.4730361159660301 RS: 4
Coherence Score:0.49627736930757965 RS: 5
Coherence Score:0.5035075494053166 RS: 6
Coherence Score:0.5009781472790498 RS: 7
Coherence Score:0.5064013010911522 RS: 8
Coherence Score:0.4907097048934452 RS: 9
Coherence Score:0.4880634164343027 RS: 10
Coherence Score:0.4679983371161482 RS: 11
Coherence Score:0.49411420608488804 RS: 12
Coherence Score:0.5105332928777638 RS: 13
Coherence Score:0.49439372940435494 RS: 14
Coherence Score:0.5007769963349165 RS: 15
Coherence Score:0.5186271692036467 RS: 16
Coherence Score:0.4866915755432076 RS: 17
Coherence Score:0.48034486664583453 RS: 18
Coherence Score:0.49445017869247776 RS: 19
Coherence Score:0.5044207620649602 RS: 20
Coherence Score:0.49327247428275084 RS: 21
Coherence Score:0.4802056547738122 RS: 22
Coherence Score:0.5075621553049546 RS: 23
Coherence Score:0.48536164118150865 RS: 24
Coherence Score:0.5027447786333153 RS: 25
Coherence Score:0.4713996327812202 RS: 26
Coherence Score:0.5061410327281279 RS: 27
Coherence Score:0.49392214884638697 RS: 28
Coherence Score:0.48027595488249886 RS: 29
```

In [81]:

```
print ("The ideal coherence score {:.4f} corresponds to random state {}".format(maxv,maxrs))
```

The ideal coherence score 0.5254 corresponds to random state 2

In [108]:

```
result_df = pd.DataFrame({'No of topics': [12], 'Best Random State ': [maxrs] , ' Coherence Score': [maxv] })
result_df.style.hide_index()
```

Out[108]:

No of topics	Best Random State	Coherence Score
12	17	0.525419

The above is the best coherence score is at random state for 12 topics lets build a model using this parameters

In [83]:

```
lda_model = LdaModel(corpus=review_bos,id2word=id2word,
                      num_topics=12,random_state=2,passes=20,iterations=15)
coherence_model_lda = CoherenceModel(model=lda_model, texts=clean_review, dictionary=id2word, coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()

print ('Coherence Score:{} RS: {}'.format(coherence_lda,2))
```

Coherence Score:0.5254187672082519 RS: 2

In [84]:

```
#Print the Keyword in the 12 topics
from pprint import pprint
pprint (lda_model.print_topics())
#The below weights include the top terms for each topic

[(0,
  '0.233*"issue" + 0.110*"network" + 0.071*"service" + 0.057*"problem" + '
  '0.053*"amazon" + 0.047*"sim" + 0.034*"video" + 0.031*"support" + '
  '0.029*"music" + 0.024*"jio"),,
(1,
  '0.245*"price" + 0.185*"feature" + 0.093*"range" + 0.069*"everything" + '
  '0.045*"super" + 0.043*"glass" + 0.034*"usage" + 0.029*"piece" + '
  '0.023*"gorilla" + 0.019*"product,"'),
(2,
  '0.105*"heat" + 0.072*"delivery" + 0.071*"update" + 0.070*"software" + '
  '0.053*"please" + 0.051*"look" + 0.049*"smartphone" + 0.046*"purchase" + '
  '0.034*"return" + 0.030*"headphone")',
(3,
  '0.221*"problem" + 0.116*"day" + 0.104*"heating" + 0.070*"screen" + '
  '0.058*"call" + 0.039*"work" + 0.037*"use" + 0.034*"option" + 0.025*"volta" '
  '+ 0.025*"week"),
(4,
  '0.264*"camera" + 0.252*"battery" + 0.123*"quality" + 0.059*"backup" + '
  '0.033*"performance" + 0.029*"life" + 0.020*"mode" + 0.016*"mark" + '
  '0.013*"card" + 0.012*"depth"),
(5,
  '0.177*"month" + 0.111*"device" + 0.084*"charge" + 0.055*"excellent" + '
  '0.053*"review" + 0.047*"hr" + 0.041*"function" + 0.039*"bit" + '
  '0.033*"power" + 0.028*"i\m"),
(6,
  '0.624*"product" + 0.100*"lenovo" + 0.055*"superb" + 0.031*"h" + '
  '0.022*"star" + 0.016*"fast" + 0.016*"till" + 0.015*"date" + 0.014*"amazon" '
  '+ 0.013*"guy"),
(7,
  '0.908*"phone" + 0.009*"game" + 0.008*"bug" + 0.005*"thing" + 0.005*"offer" '
  '+ 0.004*"smart" + 0.004*"lag" + 0.004*"k8note" + 0.004*"part" + '
  '0.004*"mob"),
(8,
  '0.103*"charger" + 0.095*"speaker" + 0.079*"phone," + 0.068*"sound" + '
  '0.065*"hai" + 0.059*"handset" + 0.053*"speed" + 0.050*"turbo" + 0.045*"ok" '
  '+ 0.027*"ho"),
(9,
  '0.639*"mobile" + 0.073*"budget" + 0.040*"expectation" + 0.034*"worth" + '
  '0.026*"signal" + 0.026*"lenovo" + 0.018*"internet" + 0.014*"headset" + '
  '0.012*"weight" + 0.011*"change"),
(10,
  '0.197*"note" + 0.135*"k8" + 0.108*"performance" + 0.072*"money" + '
  '0.049*"lenovo" + 0.047*"display" + 0.044*"value" + 0.030*"cast" + '
  '0.022*"android" + 0.021*"model"),
(11,
  '0.144*"time" + 0.088*"battery" + 0.078*"waste" + 0.056*"experience" + '
  '0.055*"hour" + 0.052*"lot" + 0.042*"drain" + 0.039*"money" + '
  '0.037*"awesome" + 0.032*"box")]

```

In [85]:

```
print('\nPerplexity: ', lda_model.log_perplexity(review_bos))
```

```
Perplexity: -5.0435807887740935
```

Visualize the 12 topics keywords and their top terms from visualization

In [86]:

```
pyLDAvis.enable_notebook()
vis = pyLDAvis.gensim.prepare(lda_model,review_bos, id2word)
vis
```

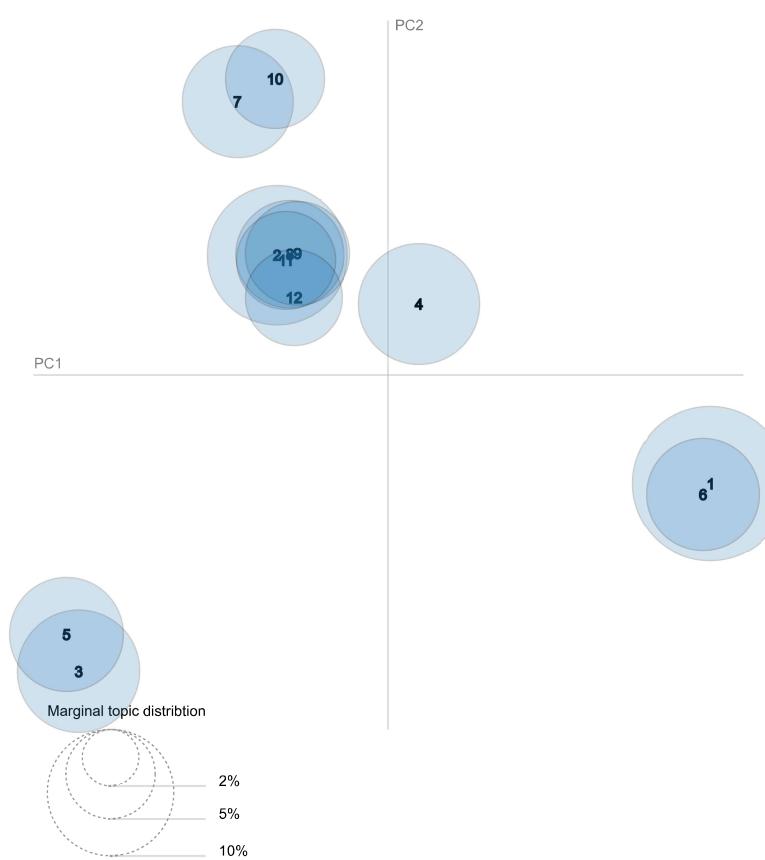
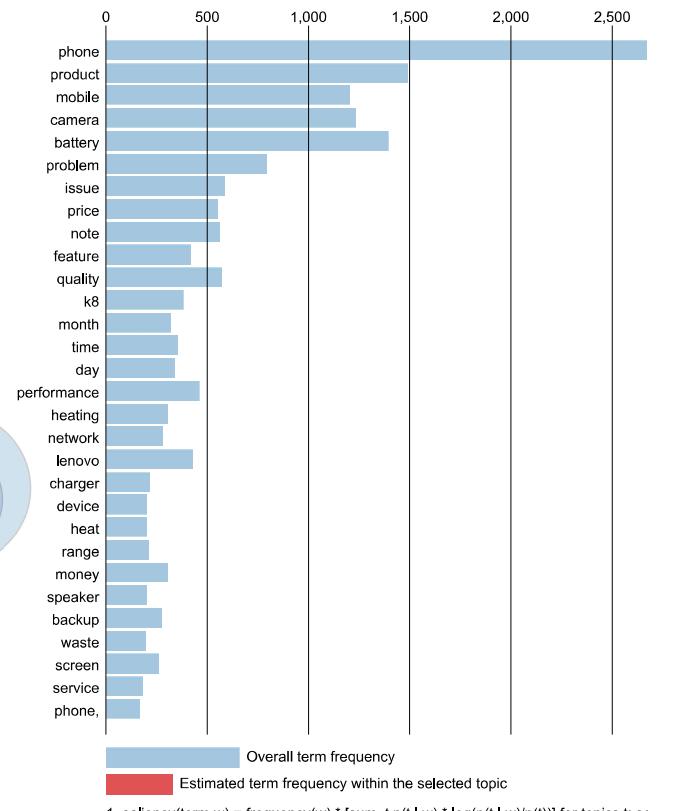
Out[86]:

Selected Topic: 0

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 1$



Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Salient Terms⁽¹⁾

Topics through business lens

Topics : '0', '1', '3' , can be combined to as they describe about HardWare Issues**

Topics : '4', '5'--> Software Issues and Connectivity

Topics : '6','12' --> doesn't shed enough light to the issues

Topics : '10','11'--> Battery Heating issue

Topics : '7','8' --> Positive talks

Creating model to find an optimal no of topics for the review token extracted

In [89]:

```
maxvs = 0
maxnt = 0
maxrs = 0

for rs in range(1,40):
    for nt in range(2,15):

        model_2 = LdaModel(corpus=review_bos,id2word=id2word,num_topics=nt,random_state=rs,
                            passes=10,iterations=20,per_word_topics=True)
        coherence_model_lda_2 = CoherenceModel(model=model_2, texts=clean_review, dictionary=id2word, coherence='c_v')
        coherence_lda_2 = coherence_model_lda_2.get_coherence()

        print('Conherence {} for RandomState : {} , No of Topics : {} '.format(coherence_lda_2,rs,nt))

    if coherence_lda_2 > maxvs:
        maxvs=coherence_lda_2
        maxrs=rs
        maxnt=nt

print('Ideal no of topics from the model is {:.4f} at Randaom State {} with C_V of :{} '.format(maxnt,maxrs,maxvs))
```

Conherence 0.23334116722785825 for RandomState : 1 , No of Topics : 2
Conherence 0.3373681713709262 for RandomState : 1 , No of Topics : 3
Conherence 0.3510231255667889 for RandomState : 1 , No of Topics : 4
Conherence 0.3881914343783445 for RandomState : 1 , No of Topics : 5
Conherence 0.391086267580487924 for RandomState : 1 , No of Topics : 6
Conherence 0.41526380737514 for RandomState : 1 , No of Topics : 7
Conherence 0.44411616488104283 for RandomState : 1 , No of Topics : 8
Conherence 0.45894481552386945 for RandomState : 1 , No of Topics : 9
Conherence 0.4707256010940135 for RandomState : 1 , No of Topics : 10
Conherence 0.4813485611454263 for RandomState : 1 , No of Topics : 11
Conherence 0.515087395054933 for RandomState : 1 , No of Topics : 12
Conherence 0.5319601291407904 for RandomState : 1 , No of Topics : 13
Conherence 0.5165064919349295 for RandomState : 1 , No of Topics : 14
Conherence 0.2715145019755621 for RandomState : 2 , No of Topics : 2
Conherence 0.33135681318576327 for RandomState : 2 , No of Topics : 3
Conherence 0.3339225259927116 for RandomState : 2 , No of Topics : 4
Conherence 0.3748426642656579 for RandomState : 2 , No of Topics : 5
Conherence 0.43809148652398383 for RandomState : 2 , No of Topics : 6
Conherence 0.43531749988389457 for RandomState : 2 , No of Topics : 7
Conherence 0.4582721895350929 for RandomState : 2 , No of Topics : 8
Conherence 0.4724668869960055 for RandomState : 2 , No of Topics : 9
Conherence 0.468563348110196 for RandomState : 2 , No of Topics : 10
Conherence 0.4794685979617698 for RandomState : 2 , No of Topics : 11
Conherence 0.5084682938363189 for RandomState : 2 , No of Topics : 12
Conherence 0.508948778549762 for RandomState : 2 , No of Topics : 13
Conherence 0.5270337118880032 for RandomState : 2 , No of Topics : 14
Conherence 0.25249985477287656 for RandomState : 3 , No of Topics : 2
Conherence 0.34940911055203455 for RandomState : 3 , No of Topics : 3
Conherence 0.3940802120535082 for RandomState : 3 , No of Topics : 4
Conherence 0.4190135650074156 for RandomState : 3 , No of Topics : 5
Conherence 0.3993394250877563 for RandomState : 3 , No of Topics : 6
Conherence 0.42598402417856296 for RandomState : 3 , No of Topics : 7
Conherence 0.4179015984794023 for RandomState : 3 , No of Topics : 8
Conherence 0.4660737113954075 for RandomState : 3 , No of Topics : 9
Conherence 0.4768782387386749 for RandomState : 3 , No of Topics : 10
Conherence 0.46814467513597025 for RandomState : 3 , No of Topics : 11
Conherence 0.4767907978288686 for RandomState : 3 , No of Topics : 12
Conherence 0.4929413164650869 for RandomState : 3 , No of Topics : 13
Conherence 0.5052904989245703 for RandomState : 3 , No of Topics : 14
Conherence 0.2709047679357227 for RandomState : 4 , No of Topics : 2
Conherence 0.31399335561686725 for RandomState : 4 , No of Topics : 3
Conherence 0.39135179562570177 for RandomState : 4 , No of Topics : 4
Conherence 0.38419645844985 for RandomState : 4 , No of Topics : 5
Conherence 0.40499420663693203 for RandomState : 4 , No of Topics : 6
Conherence 0.3841684458329624 for RandomState : 4 , No of Topics : 7
Conherence 0.4131167806138525 for RandomState : 4 , No of Topics : 8
Conherence 0.4608874019171956 for RandomState : 4 , No of Topics : 9
Conherence 0.47553167828192305 for RandomState : 4 , No of Topics : 10
Conherence 0.4633885320186042 for RandomState : 4 , No of Topics : 11
Conherence 0.4736300875535773 for RandomState : 4 , No of Topics : 12
Conherence 0.5097153326505297 for RandomState : 4 , No of Topics : 13
Conherence 0.5136611579552924 for RandomState : 4 , No of Topics : 14
Conherence 0.2503559496299558 for RandomState : 5 , No of Topics : 2
Conherence 0.32198221929016585 for RandomState : 5 , No of Topics : 3
Conherence 0.33895998456240983 for RandomState : 5 , No of Topics : 4
Conherence 0.40393011979307775 for RandomState : 5 , No of Topics : 5
Conherence 0.39943964154544037 for RandomState : 5 , No of Topics : 6
Conherence 0.43928797469447234 for RandomState : 5 , No of Topics : 7
Conherence 0.46275589340656464 for RandomState : 5 , No of Topics : 8
Conherence 0.463823702720434 for RandomState : 5 , No of Topics : 9
Conherence 0.458057992964631 for RandomState : 5 , No of Topics : 10
Conherence 0.46304647865281545 for RandomState : 5 , No of Topics : 11
Conherence 0.49923281767166694 for RandomState : 5 , No of Topics : 12
Conherence 0.5191470815739626 for RandomState : 5 , No of Topics : 13
Conherence 0.5263572506664229 for RandomState : 5 , No of Topics : 14
Conherence 0.2417334649438565 for RandomState : 6 , No of Topics : 2
Conherence 0.3480573082315946 for RandomState : 6 , No of Topics : 3
Conherence 0.3960989749738806 for RandomState : 6 , No of Topics : 4
Conherence 0.3802164605072414 for RandomState : 6 , No of Topics : 5
Conherence 0.404885597024424 for RandomState : 6 , No of Topics : 6
Conherence 0.41380472686635666 for RandomState : 6 , No of Topics : 7
Conherence 0.4151816208565916 for RandomState : 6 , No of Topics : 8
Conherence 0.44972813421044994 for RandomState : 6 , No of Topics : 9
Conherence 0.4725190672808929 for RandomState : 6 , No of Topics : 10
Conherence 0.49714439032277846 for RandomState : 6 , No of Topics : 11
Conherence 0.5057602402794533 for RandomState : 6 , No of Topics : 12
Conherence 0.5072193516146062 for RandomState : 6 , No of Topics : 13
Conherence 0.5212070863983669 for RandomState : 6 , No of Topics : 14
Conherence 0.24606761762673238 for RandomState : 7 , No of Topics : 2
Conherence 0.371809561704238 for RandomState : 7 , No of Topics : 3
Conherence 0.3621808977015266 for RandomState : 7 , No of Topics : 4
Conherence 0.41027475339084996 for RandomState : 7 , No of Topics : 5
Conherence 0.4068402345242638 for RandomState : 7 , No of Topics : 6
Conherence 0.43610537610562045 for RandomState : 7 , No of Topics : 7
Conherence 0.41106486459637775 for RandomState : 7 , No of Topics : 8
Conherence 0.4347453625969096 for RandomState : 7 , No of Topics : 9
Conherence 0.4825177795802766 for RandomState : 7 , No of Topics : 10
Conherence 0.47042868079143985 for RandomState : 7 , No of Topics : 11
Conherence 0.5012617217939536 for RandomState : 7 , No of Topics : 12
Conherence 0.5153110909753829 for RandomState : 7 , No of Topics : 13
Conherence 0.5424792888578882 for RandomState : 7 , No of Topics : 14

Conherence 0.33598299619771255 for RandomState : 29 , No of Topics : 5
Conherence 0.34305217815015987 for RandomState : 29 , No of Topics : 6
Conherence 0.40448870763490835 for RandomState : 29 , No of Topics : 7
Conherence 0.4161688149448007 for RandomState : 29 , No of Topics : 8
Conherence 0.4578599911145724 for RandomState : 29 , No of Topics : 9
Conherence 0.46967429990542914 for RandomState : 29 , No of Topics : 10
Conherence 0.4691784852803561 for RandomState : 29 , No of Topics : 11
Conherence 0.49534292848188693 for RandomState : 29 , No of Topics : 12
Conherence 0.4951035698638815 for RandomState : 29 , No of Topics : 13
Conherence 0.5029470519250568 for RandomState : 29 , No of Topics : 14
Conherence 0.22952082323948467 for RandomState : 30 , No of Topics : 2
Conherence 0.29403488493303 for RandomState : 30 , No of Topics : 3
Conherence 0.3760324340798761 for RandomState : 30 , No of Topics : 4
Conherence 0.4000301036184024 for RandomState : 30 , No of Topics : 5
Conherence 0.38252846810345725 for RandomState : 30 , No of Topics : 6
Conherence 0.43987345363337543 for RandomState : 30 , No of Topics : 7
Conherence 0.4201944545713412 for RandomState : 30 , No of Topics : 8
Conherence 0.4312027876440696 for RandomState : 30 , No of Topics : 9
Conherence 0.45698980823512797 for RandomState : 30 , No of Topics : 10
Conherence 0.4522238544460704 for RandomState : 30 , No of Topics : 11
Conherence 0.4938336599111423 for RandomState : 30 , No of Topics : 12
Conherence 0.4882590194730105 for RandomState : 30 , No of Topics : 13
Conherence 0.523123092108464 for RandomState : 30 , No of Topics : 14
Conherence 0.24317255174085778 for RandomState : 31 , No of Topics : 2
Conherence 0.3444151747765884 for RandomState : 31 , No of Topics : 3
Conherence 0.39236093989272913 for RandomState : 31 , No of Topics : 4
Conherence 0.3492891679376471 for RandomState : 31 , No of Topics : 5
Conherence 0.3890730437068452 for RandomState : 31 , No of Topics : 6
Conherence 0.43908680063150257 for RandomState : 31 , No of Topics : 7
Conherence 0.44550080310635454 for RandomState : 31 , No of Topics : 8
Conherence 0.4399906009300705 for RandomState : 31 , No of Topics : 9
Conherence 0.4670585316575508 for RandomState : 31 , No of Topics : 10
Conherence 0.4692657519014806 for RandomState : 31 , No of Topics : 11
Conherence 0.47434005942879537 for RandomState : 31 , No of Topics : 12
Conherence 0.5103771417558709 for RandomState : 31 , No of Topics : 13
Conherence 0.5085657687015592 for RandomState : 31 , No of Topics : 14
Conherence 0.24362194681623112 for RandomState : 32 , No of Topics : 2
Conherence 0.37125321845511916 for RandomState : 32 , No of Topics : 3
Conherence 0.366737967730244 for RandomState : 32 , No of Topics : 4
Conherence 0.38287857967965844 for RandomState : 32 , No of Topics : 5
Conherence 0.3992435823213408 for RandomState : 32 , No of Topics : 6
Conherence 0.4113272749340105 for RandomState : 32 , No of Topics : 7
Conherence 0.4251845992147302 for RandomState : 32 , No of Topics : 8
Conherence 0.44914841573021175 for RandomState : 32 , No of Topics : 9
Conherence 0.466979617348421 for RandomState : 32 , No of Topics : 10
Conherence 0.4779001301690716 for RandomState : 32 , No of Topics : 11
Conherence 0.5036295073452796 for RandomState : 32 , No of Topics : 12
Conherence 0.5237875283674945 for RandomState : 32 , No of Topics : 13
Conherence 0.5383362771560783 for RandomState : 32 , No of Topics : 14
Conherence 0.2718516999628181 for RandomState : 33 , No of Topics : 2
Conherence 0.40631483346435654 for RandomState : 33 , No of Topics : 3
Conherence 0.40692425979481694 for RandomState : 33 , No of Topics : 4
Conherence 0.4169801730876238 for RandomState : 33 , No of Topics : 5
Conherence 0.4043771544238964 for RandomState : 33 , No of Topics : 6
Conherence 0.4333578444882765 for RandomState : 33 , No of Topics : 7
Conherence 0.41312948691819895 for RandomState : 33 , No of Topics : 8
Conherence 0.43426910648992983 for RandomState : 33 , No of Topics : 9
Conherence 0.4664934489924121 for RandomState : 33 , No of Topics : 10
Conherence 0.5004936127179158 for RandomState : 33 , No of Topics : 11
Conherence 0.5010559712197542 for RandomState : 33 , No of Topics : 12
Conherence 0.5052444817904782 for RandomState : 33 , No of Topics : 13
Conherence 0.516134384210459 for RandomState : 33 , No of Topics : 14
Conherence 0.24271901923431027 for RandomState : 34 , No of Topics : 2
Conherence 0.3380004770223281 for RandomState : 34 , No of Topics : 3
Conherence 0.4141486352071077 for RandomState : 34 , No of Topics : 4
Conherence 0.3883021146126074 for RandomState : 34 , No of Topics : 5
Conherence 0.4107967478382694 for RandomState : 34 , No of Topics : 6
Conherence 0.4185412780972039 for RandomState : 34 , No of Topics : 7
Conherence 0.444832777967837 for RandomState : 34 , No of Topics : 8
Conherence 0.4393877390318207 for RandomState : 34 , No of Topics : 9
Conherence 0.4621510420101428 for RandomState : 34 , No of Topics : 10
Conherence 0.4911497780486293 for RandomState : 34 , No of Topics : 11
Conherence 0.5060727539240676 for RandomState : 34 , No of Topics : 12
Conherence 0.5124932880819097 for RandomState : 34 , No of Topics : 13
Conherence 0.51447539075435 for RandomState : 34 , No of Topics : 14
Conherence 0.2793299908815185 for RandomState : 35 , No of Topics : 2
Conherence 0.3003682985012284 for RandomState : 35 , No of Topics : 3
Conherence 0.40457229073628176 for RandomState : 35 , No of Topics : 4
Conherence 0.4106461525249546 for RandomState : 35 , No of Topics : 5
Conherence 0.4140767092114023 for RandomState : 35 , No of Topics : 6
Conherence 0.43816007470769464 for RandomState : 35 , No of Topics : 7
Conherence 0.4366595865792161 for RandomState : 35 , No of Topics : 8
Conherence 0.43220853450271557 for RandomState : 35 , No of Topics : 9
Conherence 0.4425899502658666 for RandomState : 35 , No of Topics : 10
Conherence 0.4713522473582385 for RandomState : 35 , No of Topics : 11
Conherence 0.49094857449791274 for RandomState : 35 , No of Topics : 12
Conherence 0.4840831025742133 for RandomState : 35 , No of Topics : 13
Conherence 0.5050332150492656 for RandomState : 35 , No of Topics : 14
Conherence 0.2583666057327388 for RandomState : 36 , No of Topics : 2
Conherence 0.354233637170645 for RandomState : 36 , No of Topics : 3
Conherence 0.37933659254163127 for RandomState : 36 , No of Topics : 4
Conherence 0.38839968990351503 for RandomState : 36 , No of Topics : 5

```

Conherence 0.3950635515071414 for RandomState : 36 , No of Topics : 6
Conherence 0.4031997074994476 for RandomState : 36 , No of Topics : 7
Conherence 0.39980796620743264 for RandomState : 36 , No of Topics : 8
Conherence 0.4206287886426397 for RandomState : 36 , No of Topics : 9
Conherence 0.4517792097796918 for RandomState : 36 , No of Topics : 10
Conherence 0.46976059109499035 for RandomState : 36 , No of Topics : 11
Conherence 0.4913617354143425 for RandomState : 36 , No of Topics : 12
Conherence 0.5085981576124662 for RandomState : 36 , No of Topics : 13
Conherence 0.5073567402604171 for RandomState : 36 , No of Topics : 14
Conherence 0.249591131506741 for RandomState : 37 , No of Topics : 2
Conherence 0.30169642326008983 for RandomState : 37 , No of Topics : 3
Conherence 0.3668468157446809 for RandomState : 37 , No of Topics : 4
Conherence 0.36618059195379093 for RandomState : 37 , No of Topics : 5
Conherence 0.35242996545976585 for RandomState : 37 , No of Topics : 6
Conherence 0.38559511232141697 for RandomState : 37 , No of Topics : 7
Conherence 0.4055797944495404 for RandomState : 37 , No of Topics : 8
Conherence 0.43617754724734775 for RandomState : 37 , No of Topics : 9
Conherence 0.4112397212439719 for RandomState : 37 , No of Topics : 10
Conherence 0.4852484574252928 for RandomState : 37 , No of Topics : 11
Conherence 0.46212842438265395 for RandomState : 37 , No of Topics : 12
Conherence 0.509553849238076 for RandomState : 37 , No of Topics : 13
Conherence 0.519531120099594 for RandomState : 37 , No of Topics : 14
Conherence 0.2894188572037323 for RandomState : 38 , No of Topics : 2
Conherence 0.3566394397867128 for RandomState : 38 , No of Topics : 3
Conherence 0.39801796344507406 for RandomState : 38 , No of Topics : 4
Conherence 0.4147856143076448 for RandomState : 38 , No of Topics : 5
Conherence 0.415564733379549 for RandomState : 38 , No of Topics : 6
Conherence 0.4264281494665715 for RandomState : 38 , No of Topics : 7
Conherence 0.4390101604858209 for RandomState : 38 , No of Topics : 8
Conherence 0.45982980037396814 for RandomState : 38 , No of Topics : 9
Conherence 0.44490396026257617 for RandomState : 38 , No of Topics : 10
Conherence 0.48565978025373274 for RandomState : 38 , No of Topics : 11
Conherence 0.5098210314344522 for RandomState : 38 , No of Topics : 12
Conherence 0.5352893238575029 for RandomState : 38 , No of Topics : 13
Conherence 0.5170764478418522 for RandomState : 38 , No of Topics : 14
Conherence 0.2616772541039823 for RandomState : 39 , No of Topics : 2
Conherence 0.3710288382967996 for RandomState : 39 , No of Topics : 3
Conherence 0.338292075140458 for RandomState : 39 , No of Topics : 4
Conherence 0.3721871521411896 for RandomState : 39 , No of Topics : 5
Conherence 0.3778405829058536 for RandomState : 39 , No of Topics : 6
Conherence 0.3604713202402921 for RandomState : 39 , No of Topics : 7
Conherence 0.3954971336655402 for RandomState : 39 , No of Topics : 8
Conherence 0.41890378853227395 for RandomState : 39 , No of Topics : 9
Conherence 0.44616986325227154 for RandomState : 39 , No of Topics : 10
Conherence 0.4864386608862043 for RandomState : 39 , No of Topics : 11
Conherence 0.4837836795663575 for RandomState : 39 , No of Topics : 12
Conherence 0.5017919092587653 for RandomState : 39 , No of Topics : 13
Conherence 0.5181922241914866 for RandomState : 39 , No of Topics : 14
Idel no of topics from the model is 14.0000 at Randaom State 7 with C_V of :0.5424792888578882

```

In [90]:

```

#Conherence 0.5319601291407904 for RandomState : 1 , No of Topics : 13
#Conherence 0.5270337118880032 for RandomState : 2 , No of Topics : 14
#Conherence 0.5263572506664229 for RandomState : 5 , No of Topics : 14
#Conherence 0.542479288578882 for RandomState : 7 , No of Topics : 14
#Conherence 0.49714439032277846 for RandomState : 6 , No of Topics : 11
#Conherence 0.48251777958802766 for RandomState : 7 , No of Topics : 10
#Conherence 0.5041377562609682 for RandomState : 16 , No of Topics : 12
#Conherence 0.4830898148101635 for RandomState : 17 , No of Topics : 9

```

The best no of topics from bove itterations according to me should be 10 or 11 , tough 13 ,14 show higher coherence, we saw that in 12 topics it self there was extarcted

In [168]:

```

LDA_model=LdaModel(corpus=review_bos,id2word=id2word,num_topics=9,
                     random_state=17,passes=15,iterations=3,per_word_topics=True)
LDA_coherence=CoherenceModel(model=LDA_model,texts=clean_review,dictionary=id2word, coherence='c_v')
coherence_score=LDA_coherence.get_coherence()

print('The Conherance C_V is {:.4f} for {} no of topics at radom state {}'.format(coherence_score,9,17))

```

The Conherance C_V is 0.4989 for 9 no of topics at radom state 17

In [169]:

```

maxrs = 17
no_t = 9
result_df = pd.DataFrame({'No of topics': [no_t], 'Best Random State ': [maxrs] , 'Coherence Score': [coherence_score]}) 
result_df.style.hide_index()

```

Out[169]:

No of topics	Best Random State	Coherence Score
9	17	0.498865

In [173]:

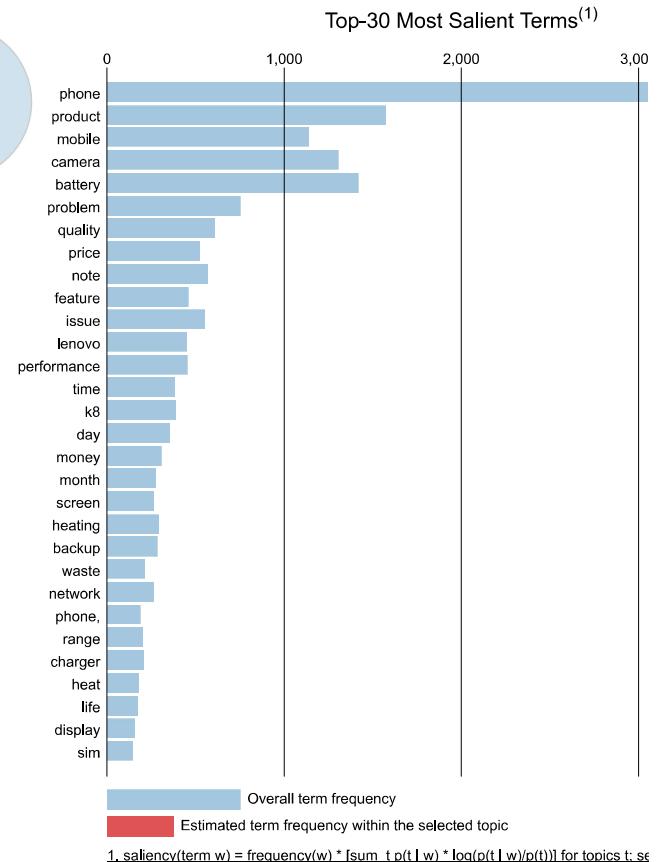
```
pyLDAvis.enable_notebook()
vis=pyLDAvis.gensim.prepare(LDA_model,review_bos,id2word)
vis
```

Out[173]:

Selected Topic: 0

Slide to adjust relevance metric:(2)
 $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



a. Name each of the identified topics.

b. Create a table with the topic name and the top 10 terms in each to present to the business

In [233]:

```
topics=LDA_model.show_topics(formatted=False)
topics
```

Out[233]:

```
[(),  
 [('battery', 0.30386594),  
 ('problem', 0.16116351),  
 ('issue', 0.11749255),  
 ('heating', 0.06226815),  
 ('backup', 0.060860317),  
 ('network', 0.055654302),  
 ('device', 0.033521615),  
 ('lot', 0.017315976),  
 ('h', 0.016107192),  
 ('bit', 0.011844863)]),  
(1,  
 [('note', 0.15092278),  
 ('lenovo', 0.1188589),  
 ('k8', 0.1023724),  
 ('day', 0.09269404),  
 ('charger', 0.053679917),  
 ('sound', 0.035410564),  
 ('experience', 0.034605768),  
 ('turbo', 0.02635747),  
 ('smartphone', 0.02184812),  
 ('charging', 0.019570775)]),  
(2,  
 [('mobile', 0.39389864),  
 ('price', 0.18176794),  
 ('screen', 0.09121415),  
 ('range', 0.06916153),  
 ('budget', 0.0451149),  
 ('headphone', 0.017634604),  
 ('photo', 0.016055191),  
 ('r', 0.01582267),  
 ('class', 0.015679348),  
 ('sale', 0.014659163)]),  
(3,  
 [('money', 0.09816713),  
 ('month', 0.08861359),  
 ('everything', 0.049513273),  
 ('charge', 0.04222779),  
 ('value', 0.041036345),  
 ('update', 0.039679546),  
 ('software', 0.039395485),  
 ('speed', 0.03450726),  
 ('please', 0.029551595),  
 ('mark', 0.027123716)]),  
(4,  
 [('camera', 0.33065346),  
 ('quality', 0.15400887),  
 ('performance', 0.11471539),  
 ('speaker', 0.044663787),  
 ('handset', 0.02783657),  
 ('work', 0.027799629),  
 ('glass', 0.022609826),  
 ('video', 0.02083782),  
 ('excellent', 0.019786445),  
 ('front', 0.01308866)]),  
(5,  
 [('product', 0.54857194),  
 ('heat', 0.06055841),  
 ('hai', 0.04305241),  
 ('delivery', 0.04147386),  
 ('super', 0.032930367),  
 ('ho', 0.018429225),  
 ('hang', 0.016670942),  
 ('set', 0.016039412),  
 ('product', 0.013525454),  
 ('k', 0.012817558)]),  
(6,  
 [('phone', 0.69813913),  
 ('call', 0.029154861),  
 ('amazon', 0.028377686),  
 ('service', 0.024743985),  
 ('superb', 0.024327762),  
 ('cast', 0.014407555),  
 ('look', 0.014397021),  
 ('support', 0.013279517),  
 ('purchase', 0.012926879),  
 ('apps', 0.009029814)]),  
(7,  
 [('feature', 0.16116065),  
 ('time', 0.13557854),  
 ('waste', 0.07419174),  
 ('life', 0.06040488),  
 ('display', 0.05442789),  
 ('hour', 0.052103817),  
 ('use', 0.043256577),  
 ('mode', 0.04154439),  
 ('drain', 0.039524868),  
 ('box', 0.02982298)]),  
(8,  
 [('phone', 0.0780598),
```

```
('sim', 0.05996975),
('option', 0.050477065),
('awesome', 0.044704728),
('ok', 0.04447702),
('processor', 0.041592702),
('thing', 0.037718635),
('music', 0.037248954),
('week', 0.036918957),
('card', 0.034489874)])]
```

In [228]:

```
issues=['Battery performance', 'Features and Details', 'Value for Money', 'Software', 'Hardware', 'Misc', 'Amazon', 'Overall Performance', 'latest
```

In [230]:

```
top_terms=[]
topic_id=[1+t[0] for t in topics]
for topic in topics:
    top_terms.append([term[0] for term in topic[1]])

topic_df = pd.DataFrame({'Topic ID': topic_id, 'Top Terms':top_terms})
```

In [231]:

```
topic_df.insert(1, 'Issues', issues)
```

In [232]:

```
topic_df
topic_df.style.hide_index()
```

Out[232]:

Topic ID	Issues	Top Terms
1	Battery performance	['battery', 'problem', 'issue', 'heating', 'backup', 'network', 'device', 'lot', 'h', 'bit']
2	Features and Details	['note', 'lenovo', 'k8', 'day', 'charger', 'sound', 'experience', 'turbo', 'smartphone', 'charging']
3	Value for Money	['mobile', 'price', 'screen', 'range', 'budget', 'headphone', 'photo', 'r', 'class', 'sale']
4	Software	['money', 'month', 'everything', 'charge', 'value', 'update', 'software', 'speed', 'please', 'mark']
5	Hardware	['camera', 'quality', 'performance', 'speaker', 'handset', 'work', 'glass', 'video', 'excellent', 'front']
6	Misc	['product', 'heat', 'hai', 'delivery', 'super', 'ho', 'hang', 'set', 'product', 'k']
7	Amazon	['phone', 'call', 'amazon', 'service', 'superb', 'cast', 'look', 'support', 'purchase', 'apps']
8	Overall Performance	['feature', 'time', 'waste', 'life', 'display', 'hour', 'use', 'mode', 'drain', 'box']
9	latest Tech support	['phone', 'sim', 'option', 'awesome', 'ok', 'processor', 'thing', 'music', 'week', 'card']

Conclusion:

POS Tagging , topic modelling using LDA and topic interpretation has been successfully completed.

LDA optimal model with 9 topics and their top 10 relevant terms have been extracted

In []: