Telecom Service That Uses Intelligent Assistant For Making All Its Agents Experts from day one

Background Story

Agents often have to handle multiple chats at the same time during rush hours. This makes the task of satisfying the customers' needs and closing the sale considerably harder since the agents have to divide their time and attention across multiple visitors.

Questions

I need to analyze chat data to understand:

1. Why and how often are customers chatting in?
2. What types of products are customers trying to buy or upgrade?
3. How often are sales made in the dataset?
4. Are there any trends from the agent messages leading up to sales?
5. What are some interesting insights from the conversations.

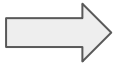
Loading and Cleaning Data

Before I perform any analysis on data, I need to load and transform it in the right format so computer can understand it.

- Transform json file into pandas dataframe with separate row for every message.
- Normalizing strings. For example: am,are, is → be, playing, played → play.
- Removing punctuations, digits, upper cases.
- Removing stop words such as the, is, at, which, and on.
- Tokenizing words (divides a text into a list of words)

Data

	messages
0	[{'text': 'I want to ask for a favor.', 'speak...
1	[{'text': 'Chat Recording: Bot: Hello! I am Ol...
2	[{'text': 'Fast connection configuration', 'sp...
3	[{'text': 'I want to ask for a favor.', 'speak...
4	[{'text': 'How can I complete my self-activati...



	chat_number	speaker_role	text	timestamp
0	message_board_0	visitor	I want to ask for a favor.	1.602636e+09
1	message_board_0	agent	Okay, just a quick question to clarify what yo...	1.602636e+09
2	message_board_0	visitor	I'm not Cresta's client yet.	1.602636e+09
3	message_board_0	agent	Give me a moment while I find a live agent for...	1.602636e+09
4	message_board_0	agent	Hi, my name is Jeri and I'm happy to help you.	1.602636e+09

Topic modeling

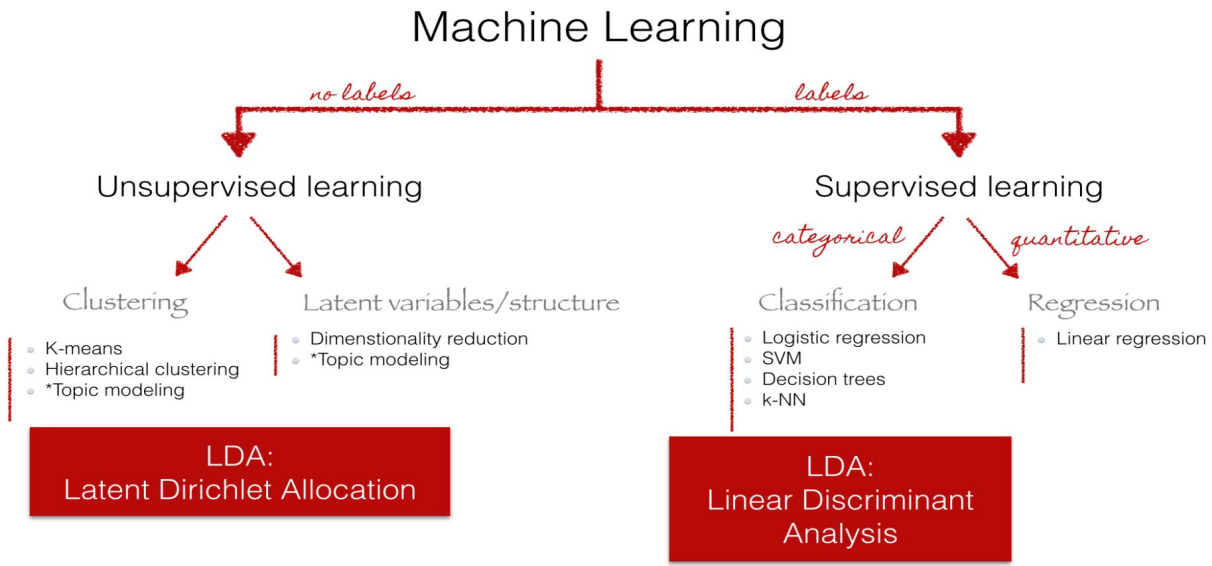
Definitions

- A topic model is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents [1]
- Topic models are a suite of algorithms that uncover the hidden thematic structure in document collections. These algorithms help us develop new ways to search, browse and summarize large archives of texts [2]
- Topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together[3]

---

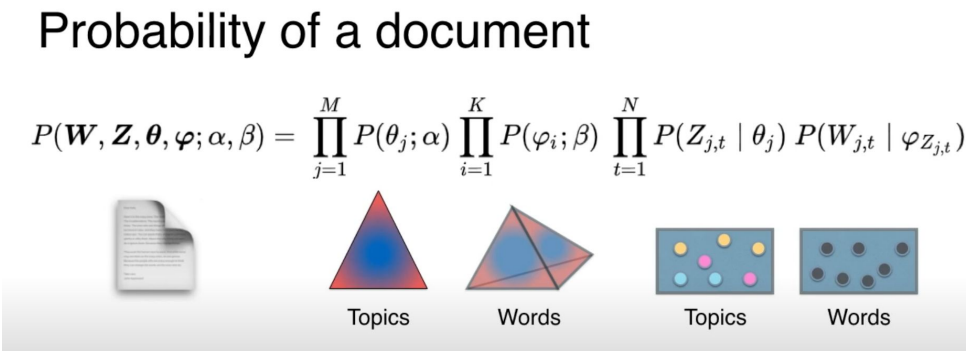
http://en.wikipedia.org/wiki/Topic_model
http://www.cs.princeton.edu/~blei/topicmodeling.html
http://mallet.cs.umass.edu/topics.php

LDA



In LDA, each document may be viewed as a mixture of various topics where each document is considered to have a set of topics that are assigned to it via LDA.

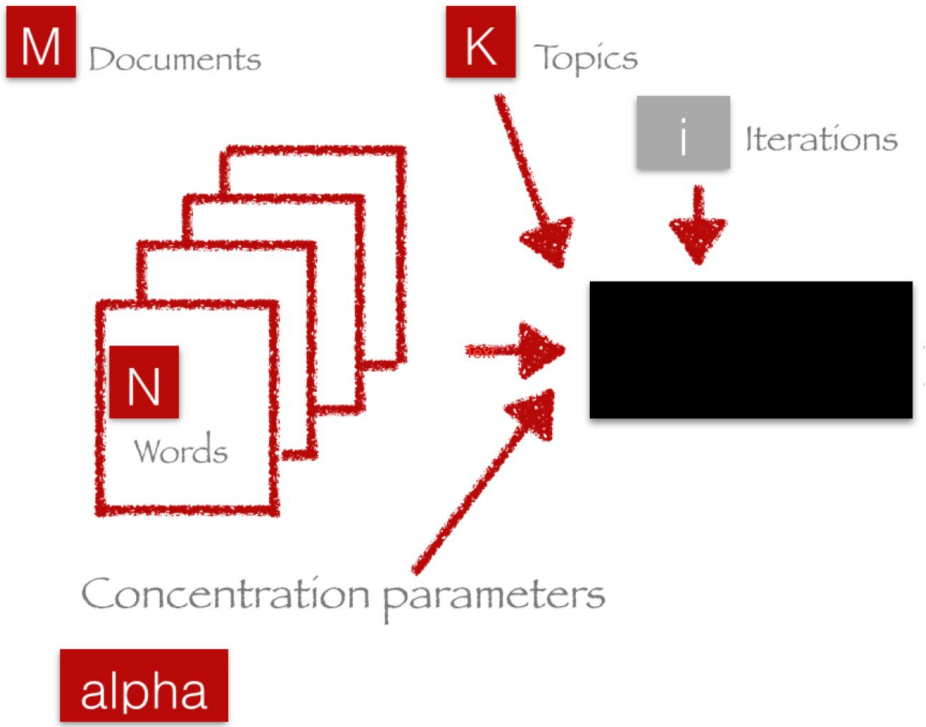
LDA



I agree this formula looks scary. But practically here, we are calculating the probability of the article being in the topics:

1. The topic distribution of the document.
2. The word distribution in the document.
3. The assignment of the topic.
4. The jth word int-the document.

LDA



Let's look at my parameters:
Corpus
Topics: 5
Alpha: 0.1* n_topics
Iterations: 50
Passes: 4

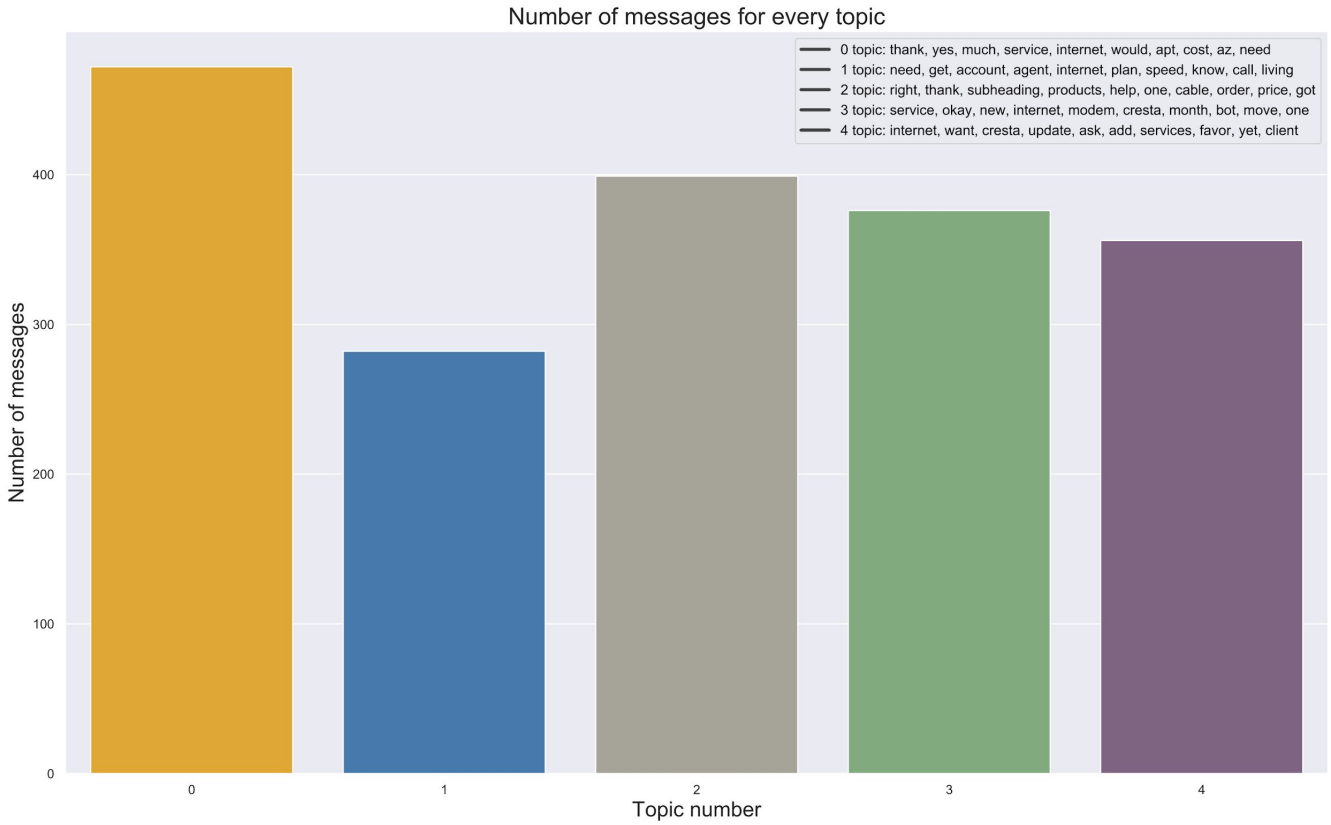
Corpus - documents.
Alpha - Dirichlet-prior concentration parameter of the per-document topic distribution.
Passes - how many times the algorithm is supposed to pass over the whole corpus.
Iterations - it controls how often we repeat a particular loop over each document.

Application of Topic Modeling on Customer's Chat

Since I don't have time to read every message, I will apply the LDA model to create 5 different topics based on my data and classify every row with the topic.

```
0 topic: thank, yes, much, service, internet, would, apt, cost, az, need
1 topic: need, get, account, agent, internet, plan, speed, know, call, living
2 topic: right, thank, subheading, products, help, one, cable, order, price, got
3 topic: service, okay, new, internet, modem, cresta, month, bot, move, one
4 topic: internet, want, cresta, update, ask, add, services, favor, yet, client
```

Application of Topic Modeling on Customer's Chat

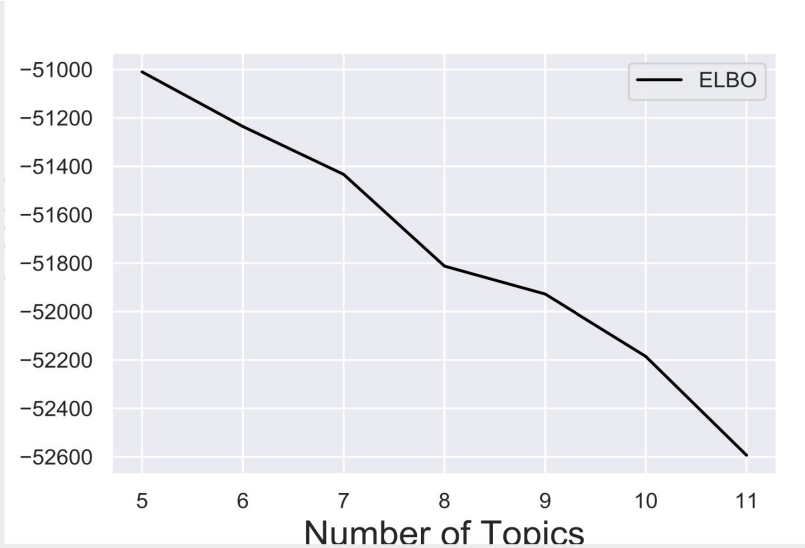


Evaluation - ELBO (evidence lower bound)

ELBO = -51010.2259

ELBO

ELBO (evidence lower bound) - is always negative, the more samples the lower the value is. The goal is to maximize ELBO score.



Evaluation - variational lower bound(ELBO)

$$
\begin{aligned}
\log p(X) &= \log \int_Z p(X, Z) \\
&= \log \int_Z p(X, Z) \frac{q(Z)}{q(Z)} \\
&= \log \left(\mathbb{E}_q \left[\frac{p(X, Z)}{q(Z)} \right] \right) \\
&\geq \mathbb{E}_q \left[\log \frac{p(X, Z)}{q(Z)} \right] \\
&= \mathbb{E}_q \left[\log p(X, Z) \right] + H[Z]
\end{aligned}
$$

- X: observed value
- Z: parameters
- q(Z): distribution of parameters
- Equation (4) applies the Jensen's inequality f (E[X]) ≤ E [f(X)] for the concave log function.

More Insights

```
Words in the library: 1654

Top 20 words:  [('right', 165), ('internet', 162), ('thank', 120), ('service', 102), ('want',
96), ('cresta', 95), ('need', 92), ('modem', 73), ('yes', 61), ('new', 58), ('ask', 55), ('ac
count', 54), ('one', 53), ('update', 52), ('services', 51), ('get', 51), ('plan', 48), ('woul
d', 46), ('okay', 42), ('month', 42)]

Words with higher occurrence: ['right', 'internet', 'thank', 'service', 'want', 'cresta', 'ne
ed', 'modem', 'yes', 'new', 'ask', 'account', 'one', 'update', 'services', 'get', 'plan', 'wo
uld', 'month', 'okay', 'cable', 'know', 'help', 'ca', 'add', 'tv', 'agent', 'much', 'lookin
g', 'see', 'think', 'move', 'favor', 'thanks', 'bot', 'going', 'number', 'speed', 'yet', 'tim
e']

Words with lower occurrence:  ['0a', 'mi424wr', 'michele', 'michelson', 'might', 'min', 'mini
cajas', 'minimize', 'minimum', 'mira', 'misokie', 'missed', 'missing', 'missokia', 'mistake
s', 'method', 'modalities', 'molinero', 'mom', 'monitored']
```

Why and how often are customers chatting in?

Coming back to the first question: Why and how often are customers chatting in? For example, they could be chatting into buy a new product or to get support on an existing service.

Based on the topics, customers are chatting to:

- **Topic 0: thank, yes, much, service, internet, would, apt, cost, az, need** - request for an appointment for a service
- **Topic 1: need, get, account, agent, internet, plan, speed, know, call, living** - get a plan with different speed
- **Topic 2: right, thank, subheading, products, help, one, cable, order, price, got -** help with pricing/ products
- **Topic 3: service, okay, new, internet, modem, cresta, month, bot, move, one -** new internet / modem
- **Topic 4: internet, want, cresta, update, ask, add, services, favor, yet, client -** update/ add a service

What types of products are customers trying to buy or upgrade?

Based on word count from the data, customers are tying to buy or upgrade the following products:

```
Most common words [('right', 165), ('internet', 162), ('thank', 120), ('service', 102), ('wan
t', 96), ('cresta', 95), ('need', 92), ('modem', 73), ('yes', 61), ('new', 58), ('ask', 55),
('account', 54), ('one', 53), ('update', 52), ('services', 51), ('get', 51), ('plan', 48),
('would', 46), ('okay', 42), ('month', 42), ('know', 40), ('cable', 40), ('help', 38), ('ca',
38), ('add', 38), ('tv', 37), ('agent', 36), ('much', 34), ('think', 32), ('see', 32), ('look
ing', 32), ('move', 31), ('favor', 30), ('thanks', 30), ('bot', 29), ('number', 29), ('goin
g', 29), ('speed', 29), ('client', 28), ('yet', 28), ('please', 28), ('time', 28), ('addres
s', 28), ('e', 28), ('order', 26), ('living', 25), ('cost', 25), ('name', 24), ('said', 24),
('could', 23), ('work', 23), ('package', 23), ('connection', 22), ('c', 22), ('go', 21), ('tw
o', 21), ('products', 21), ('still', 21), ('subheading', 21), ('last', 20), ('already', 20),
('moved', 20), ('house', 20), ('devices', 20), ('problem', 20), ('question', 19), ('support',
19), ('today', 19), ('call', 19), ('connect', 19), ('price', 19), ('something', 19), ('old',
19), ('good', 19), ('apt', 18), ('yeah', 18), ('home', 18), ('transfer', 18), ('take', 17),
('write', 17), ('complete', 17), ('activation', 17), ('az', 17), ('wifi', 17), ('trying', 1
7), ('television', 17), ('months', 17), ('got', 17), ('installation', 17), ('currently', 17),
('use', 16), ('sorry', 16), ('come', 16), ('place', 16), ('change', 16), ('tomorrow', 16),
('make', 15), ('activate', 15), ('apartment', 15), ('says', 15)]
```

How often are sales made in the dataset?

The dataset included **99** chats, **28** of the chats led to sale. Every time sale is made, agent has to send a confirmation number. By counting messages containing those two words, I was able to identify it. Next step, was to identify if it is consistent across different products. As you can see it is consistent among the following products:

```
Most common words [('order', 266), ('cresta', 117), ('help', 108), ('services', 102), ('nee
d', 86), ('internet', 83), ('today', 81), ('service', 80), ('please', 77), ('number', 68),
('account', 58), ('installation', 55), ('confirmation', 52), ('receive', 52), ('click', 51),
('new', 50), ('would', 50), ('thank', 48), ('moment', 46), ('provide', 46), ('information', 4
6), ('find', 43), ('see', 43), ('update', 42), ('looking', 41), ('speed', 41), ('get', 41),
('address', 40), ('chat', 39), ('successfully', 39), ('welcome', 39), ('sent', 39), ('happy',
38), ('month', 38), ('right', 38), ('know', 37), ('modem', 37), ('indicate', 36), ('start', 3
6), ('questions', 36), ('make', 35), ('online', 35), ('e', 35), ('add', 34), ('name', 34),
('free', 34), ('link', 33), ('customer', 31), ('streaming', 31), ('plan', 31), ('give', 30),
('want', 30), ('let', 29), ('check', 29), ('live', 28), ('everything', 28), ('home', 28), ('t
v', 28), ('agent', 27), ('email', 27), ('ask', 26), ('download', 26), ('per', 26), ('offer',
26), ('us', 26), ('sure', 25), ('mails', 25), ('includes', 25), ('page', 25), ('going', 24),
('devices', 24), ('perfect', 24), ('connect', 23), ('channels', 23), ('okay', 22), ('packag
e', 22), ('preferred', 22), ('available', 22), ('store', 21), ('gigablast', 21), ('many', 2
1), ('best', 21), ('yes', 20), ('full', 20), ('hours', 20), ('second', 20), ('processed', 2
0), ('enjoying', 20), ('hd', 20), ('mbps', 20), ('call', 19), ('question', 18), ('use', 18),
('months', 18), ('prices', 18), ('well', 17), ('take', 17), ('log', 17), ('wifi', 17), ('equi
pment', 17)]
```

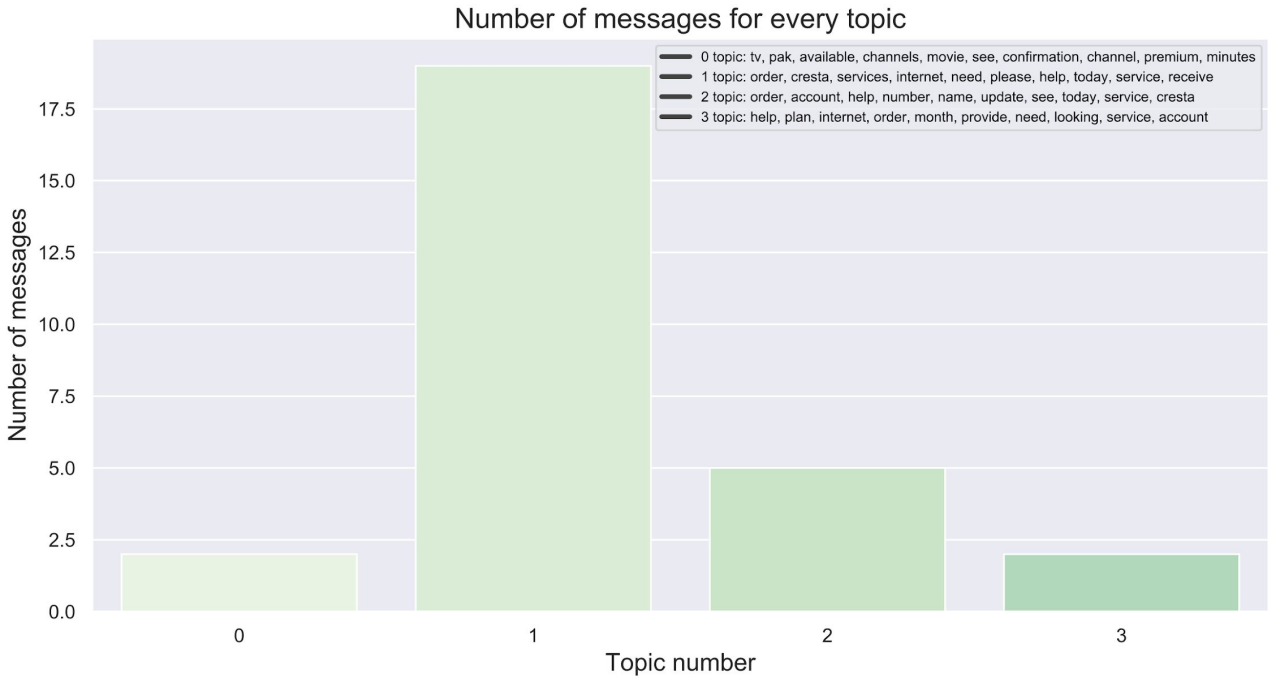
Are there any trends from the agent messages leading up to sales?

In order to understand trends with messages leading up to sales, I filtered out data by sales messages and applied the original LDA model to perform topic modeling, but this time only with 4 topics.

```
0 topic: tv, pak, available, channels, movie, see, confirmation, channel, premium, minutes
1 topic: order, cresta, services, internet, need, please, help, today, service, receive
2 topic: order, account, help, number, name, update, see, today, service, cresta
3 topic: help, plan, internet, order, month, provide, need, looking, service, account
```

Are there any trends from the agent messages leading up to sales?

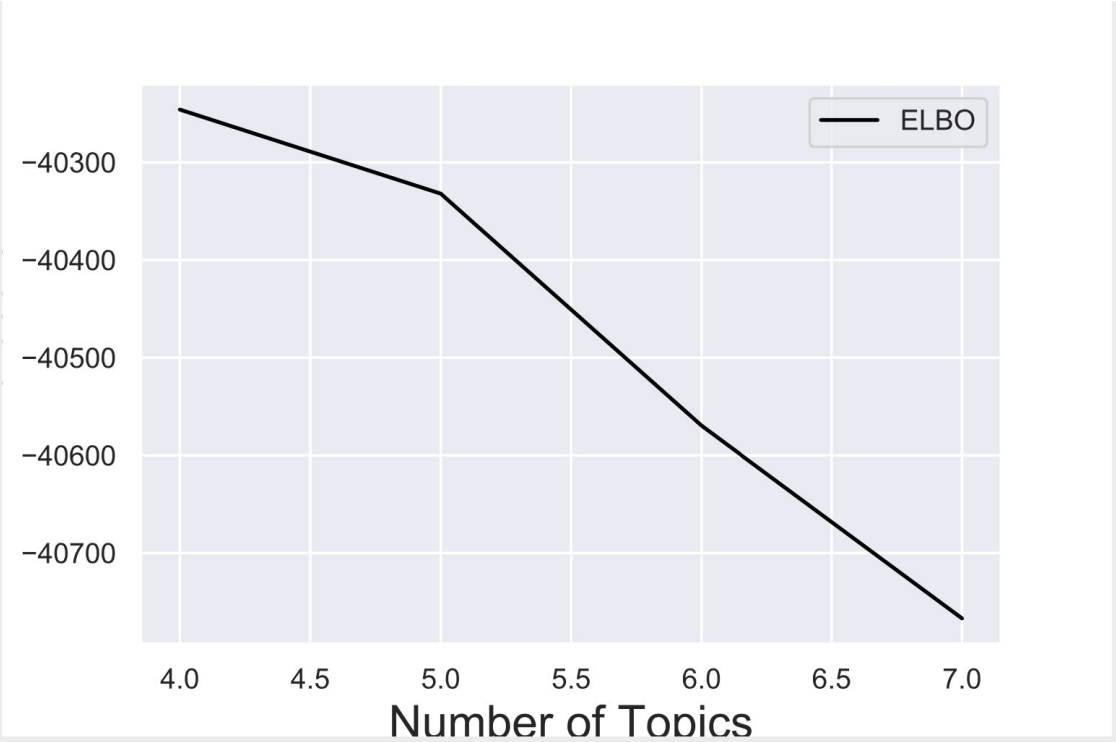
Applying topics to classify every sales message in the dataset



Evaluation - ELBO (evidence lower bound)

ELBO = -40245.2259

Evaluation - ELBO (evidence lower bound)



Are there any trends from the agent messages leading up to sales

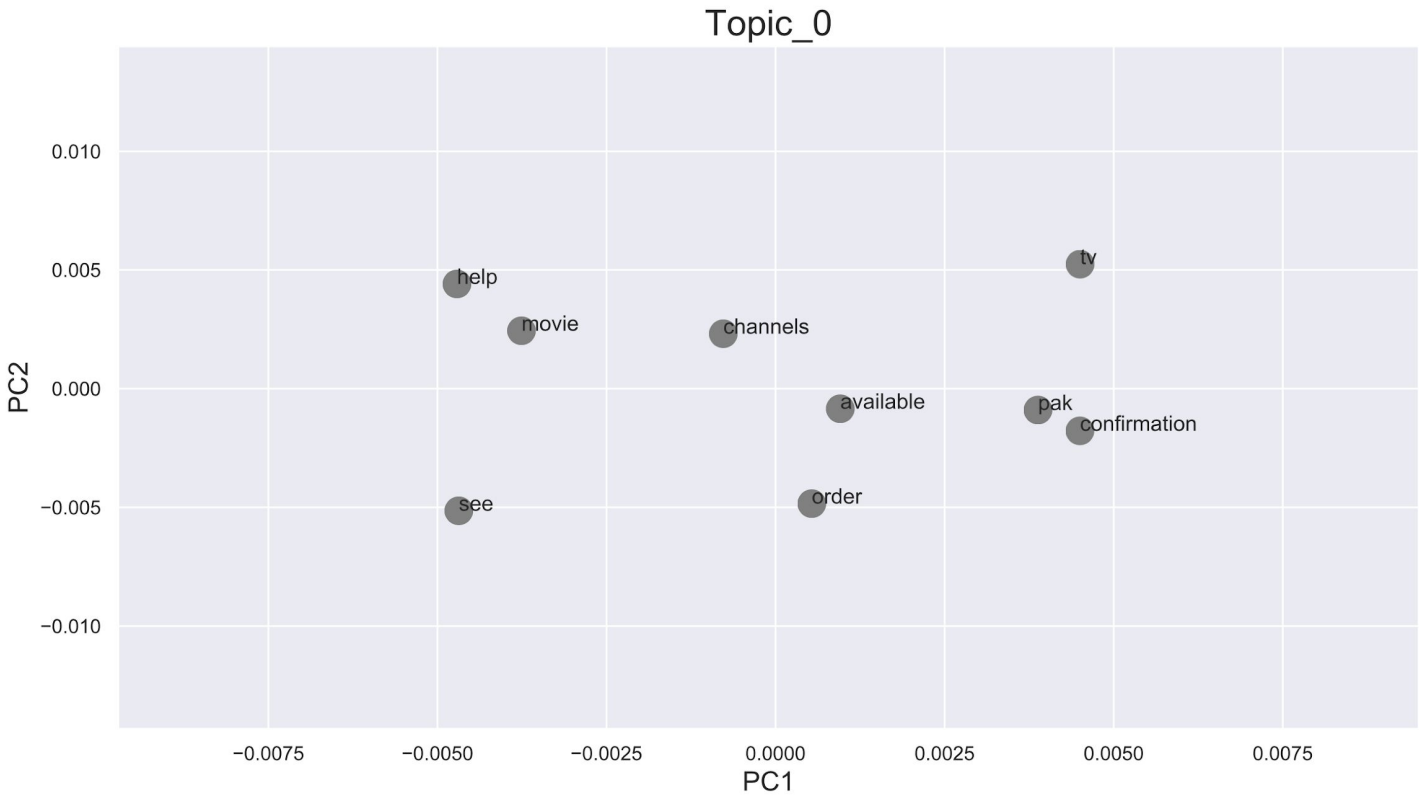
So what can we tell based on those topics:

- **0 topic: tv, pak, available, channels, movie, see, confirmation, channel, premium, minutes** - Two sales were connected to information about tv channels, premium plans; Having this information might help a sales person.
- **1 topic: order, cresta, services, internet, need, please, help, today, service, receive** - 19 sales were made with chats under topic one; This could indicate that recivng service in the same day play a big role.
- **2 topic: order, account, help, number, name, update, see, today, service, cresta** - five sales were made classified by topic two; This topic is similar to topic one, but those sales could be connected to updating of current service.
- **3 topic: help, plan, internet, order, month, provide, need, looking, service, account** - 2 sales were classified under topic 3; This could indicate that having information about different plans and internet can be helpful.

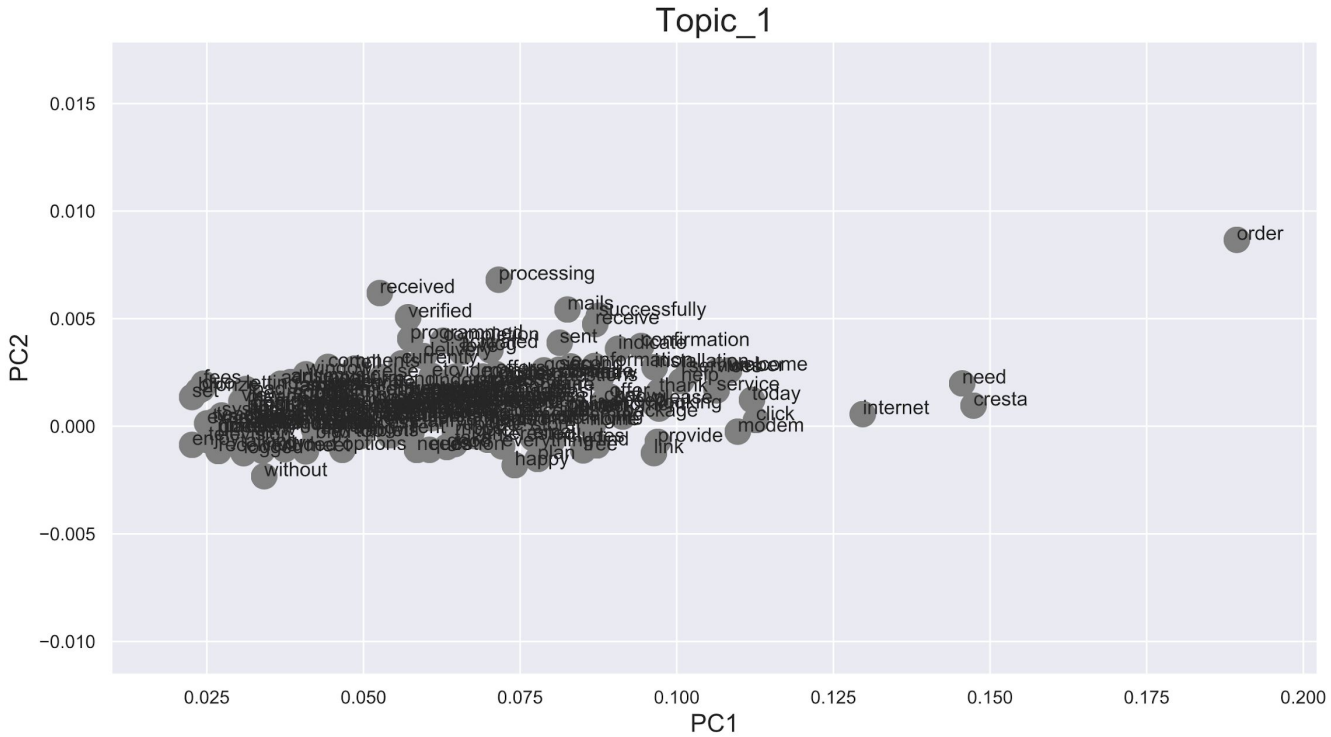
Are there any trends from the agent messages leading up to sales?

```
Words in the library: 903

Top 20 words:  [('order', 266), ('cresta', 117), ('help', 108), ('services', 102), ('need', 8
6), ('internet', 83), ('today', 81), ('service', 80), ('please', 77), ('number', 68), ('accou
nt', 58), ('installation', 55), ('confirmation', 52), ('receive', 52), ('click', 51), ('new',
50), ('would', 50), ('thank', 48), ('moment', 46), ('provide', 46)]

Words with higher occurrence: ['order', 'cresta', 'help', 'services', 'need', 'internet', 'to
day', 'service', 'please', 'number', 'account', 'installation', 'receive', 'confirmation', 'c
lick', 'new', 'would', 'thank', 'information', 'moment', 'provide', 'see', 'find', 'update',
'speed', 'get', 'looking', 'address', 'sent', 'successfully', 'chat', 'welcome', 'right', 'ha
ppy', 'month', 'know', 'modem', 'start', 'indicate', 'questions']

Words with lower occurrence:  ['10h', 'move', 'montth', 'money', 'moments', 'modify', 'moder
n', 'model', 'miss', 'minimum', 'minimize', 'messaging', 'menu', 'meets', 'makes', 'main', 'm
ade', 'lowest', 'lose', 'lori']
```

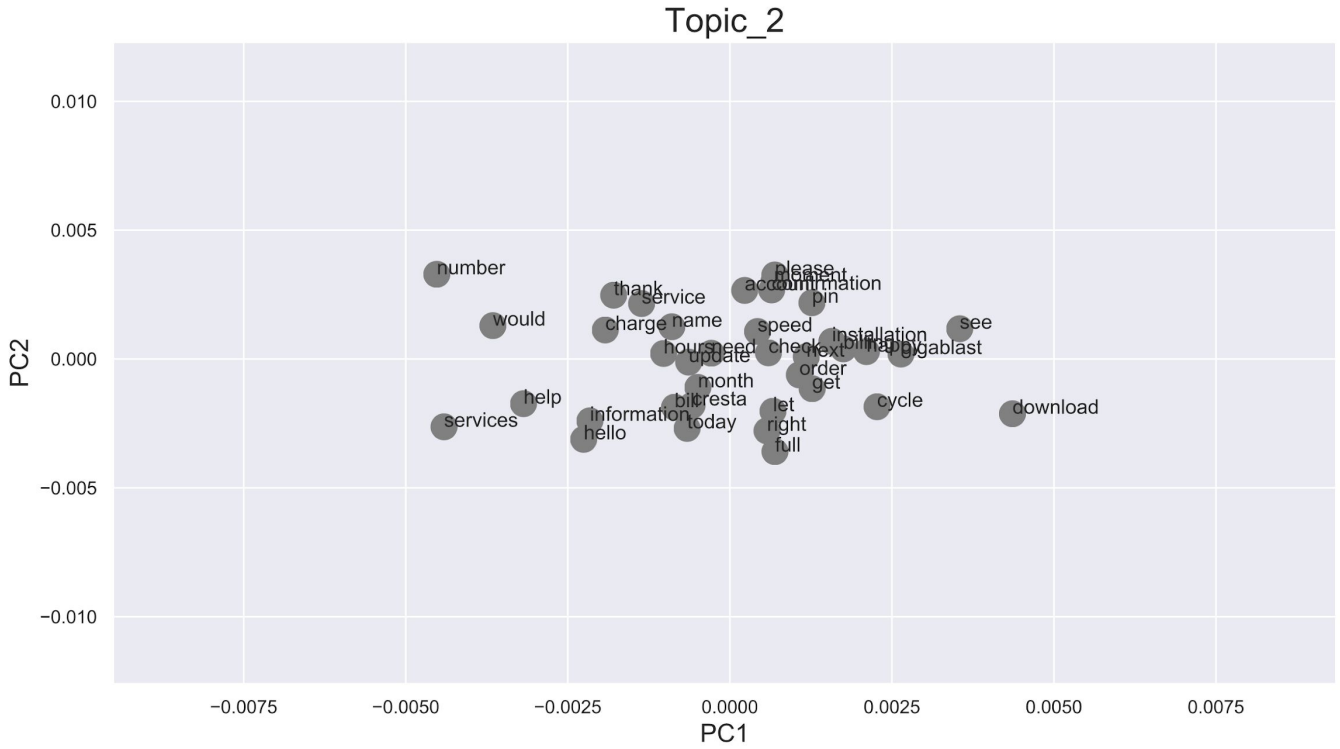
Topic 0: Tv, Pak, Available, Channels, Movie, See, Confirmation, Channel, Premium, Minutes



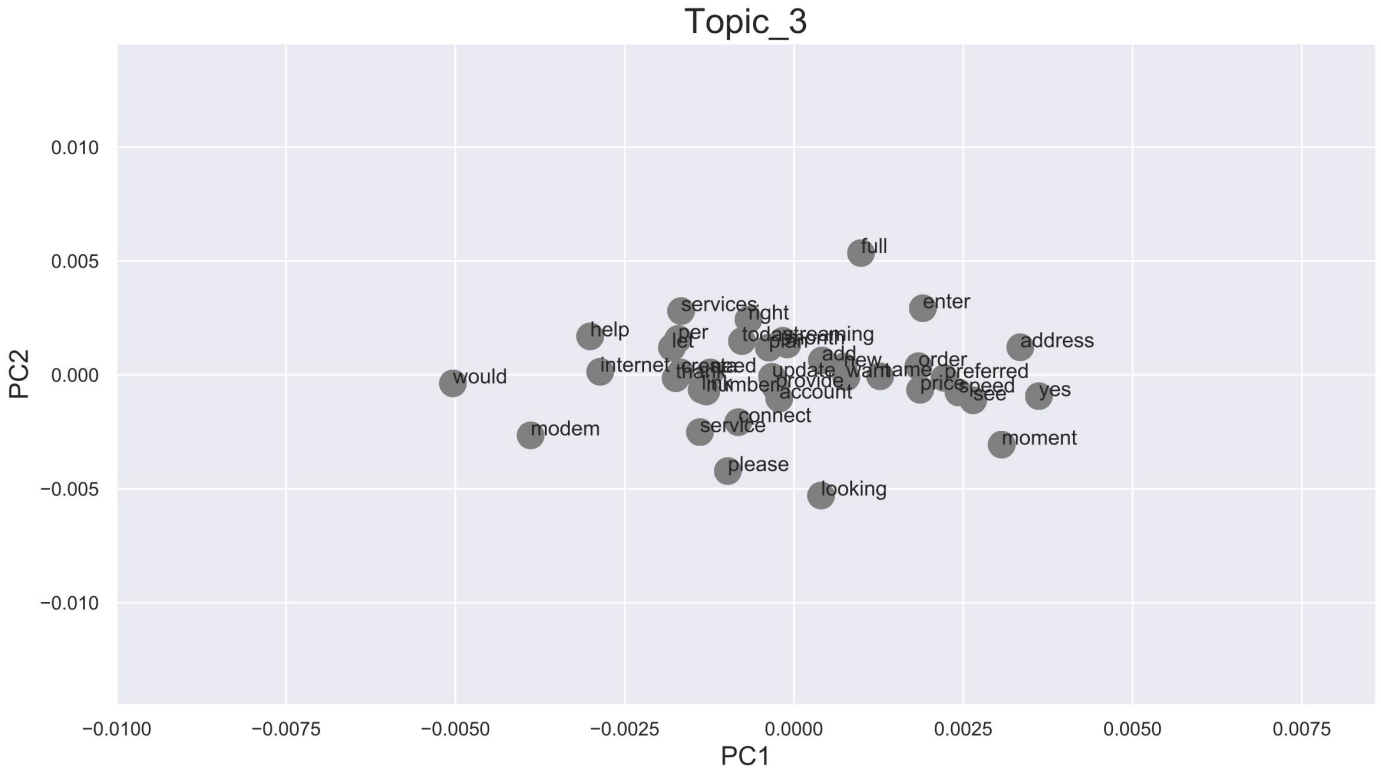
Topic 1 : Order, Cresta, Services, Internet, Need, Please, Help, Today, Service, Receive



Topic 2: Order, Account, Help, Number, Name, Update, See, Today, Service, Cresta



Topic 3: Help, Plan, Internet, Order, Month, Provide, Need, Looking, Service, Account



Summary - Customers

Based on the analysis, a few things can be strategized:

- Understand the demand of the customer:
    1. The main product is the internet; the needs are: new internet, update a plan, better speed, better price, installation.
    2. Additional products: television, cable, modem.
- Understand the frequency of the demand:
    1. Internet request (472 separate messages)
    2. Get a plan with different speed (282 individual messages)
    3. Help with pricing/products (399 individual messages)
    4. New internet/modem (376 individual messages)
    5. Update/ add a service (256 individual messages)

Summary - Sales

After a more in-depth analysis of chats that led to sales, I can infer a few things:

- The word “today” has a significant frequency. It is essential to provide service on the same day.
- Having data on different plans, installations, accounts can be helpful.
- Words like “help,” “please,” “moment,” “happy,” “welcome,” “provide,” “installation” have a significant frequency too. Using those words during conversation can be a good strategy also.

What is next

- Analyze chats that did not lead to a sale and compare them to messages that led to sales.
- Create a supervised classification model and apply it to messages to predict sales/ non-sales chats.
- Create a sentiment analysis on messages to understand what sentiment more frequently leads sales.