Natural Language Processing

Improving call centre performance

Sonja Tilly
April 2019

Natural Language Processing

Improving call centre performance

Sonja Tilly
April 2019

Project brief

- Firm XYZ has call transcripts and is looking to extract information that will add value to the business, in particular in the following areas:
    - Call centre productivity;
    - Online customer interface;
    - Product offering.

Areas of investigation

- What type of call are the agents spending most time on?
- Is there seasonality in call volume?
- What are the key topics?
- Can areas for product enhancement be identified?

Approach

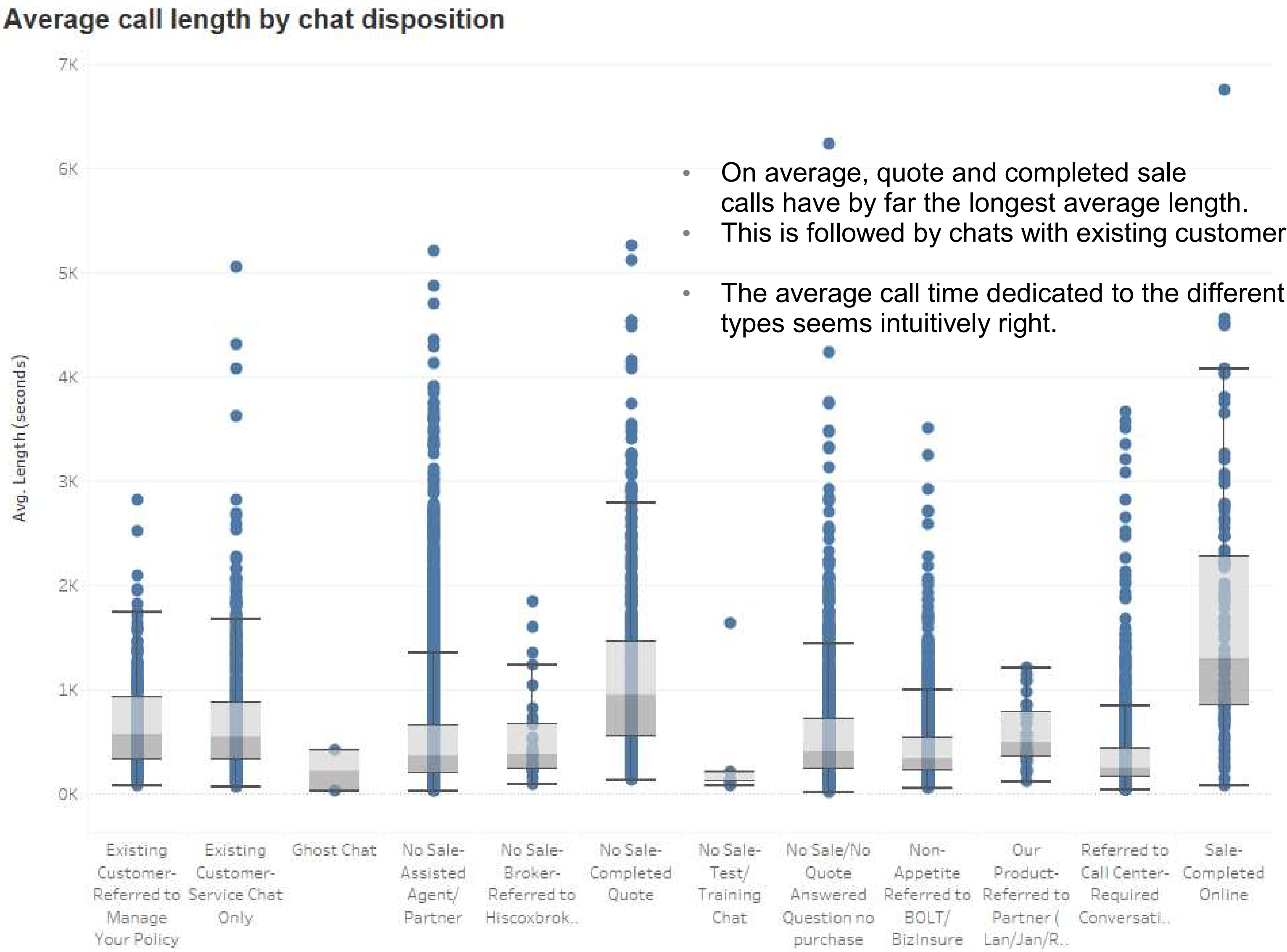
- Introduction to the dataset
- Tools used: Python libraries and Tableau
- Data exploration
- Pre-processing
- N-gram extraction
- Topic modelling
- Phrase matching
- Classification
- Conclusions & recommendations

Dataset

- 12,071 call transcripts containing unstructured text.
- Calls are classified according to chat disposition:
  - “No Sale - Assisted Agent/Partner” calls account for around 50% of all calls.
  - “No Sale – Completed Quote” and “Sale – Completed Online” only account for 4.7% and 0.9% of all calls.
- The dataset contains the start time and date of the call as well as the call length.

type
contract
else
anything
Details
Unfortunately
need
today
partner
like
able
great
may
works
Vickie
coverage
can
speaking
help
brokerage
offer
found
insurance
assist
refer
workday
thank
Hello
calling
exterior
However
work
online

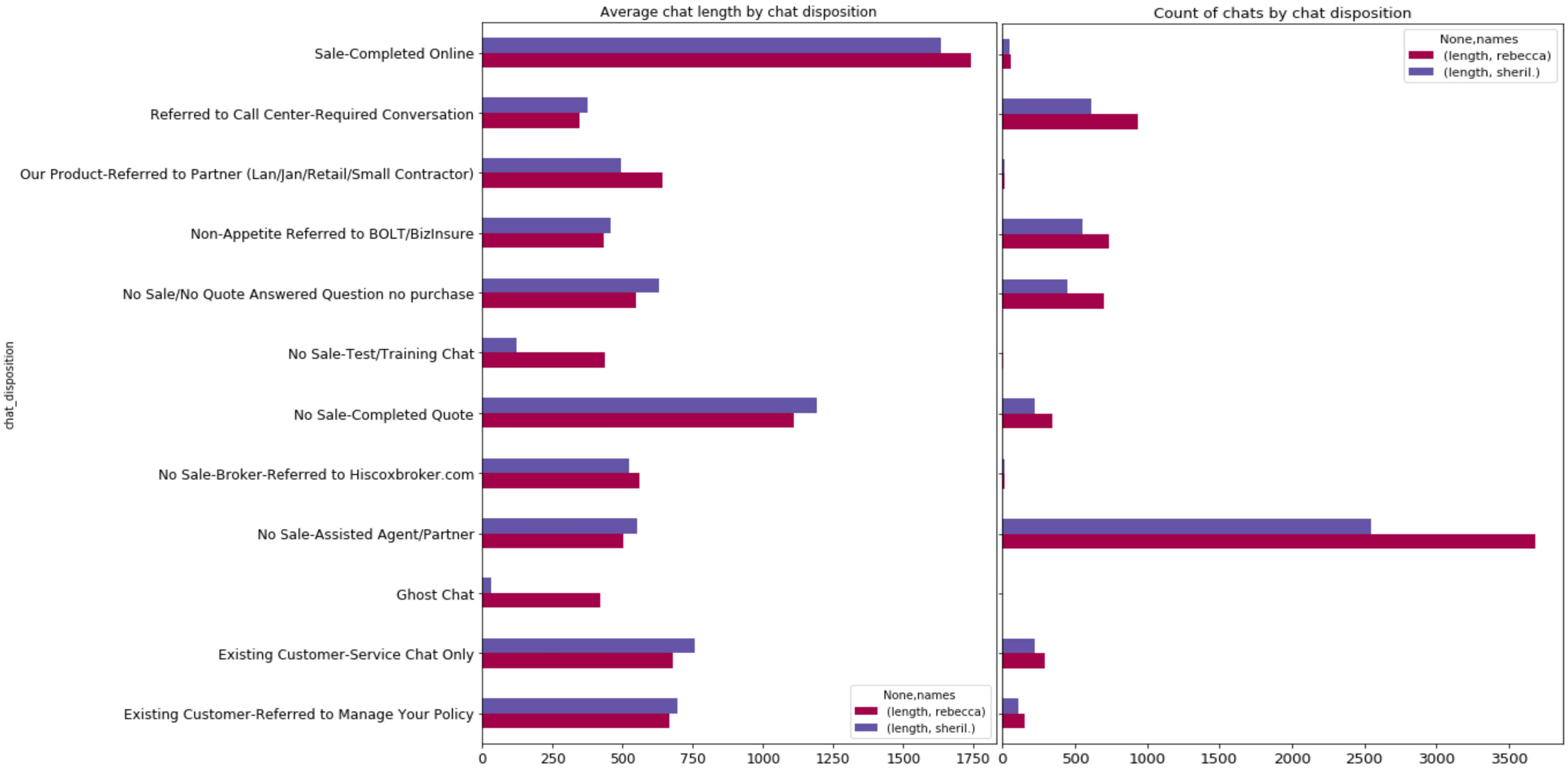
Data exploration – exploring average chat length



- On average, quote and completed sale calls have by far the longest average length.
- This is followed by chats with existing customers.

- The average call time dedicated to the different call types seems intuitively right.

Data exploration – exploring call time by agent

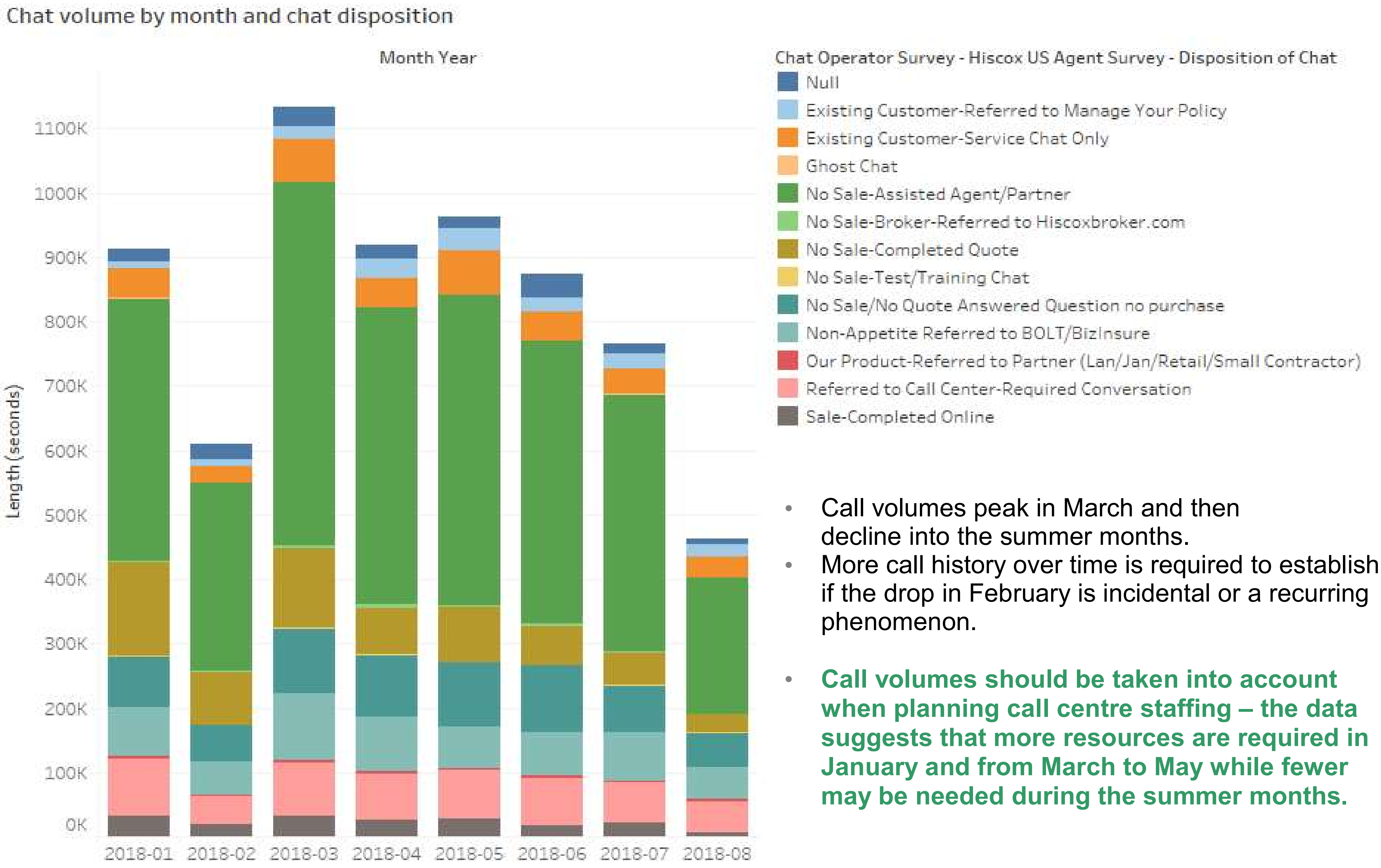


- Sheril's average chat time is longer in 7 out of 12 chat categories.
- Sheril has significantly lower count in all chat disposition categories.

Suggestion:

- **Confirm the amount of hours worked for both agents and adjust count of chats if necessary.**
- **The above indicates that Rebecca is the more productive of the two agents assuming same working hours.**

Data exploration – exploring call volume by month

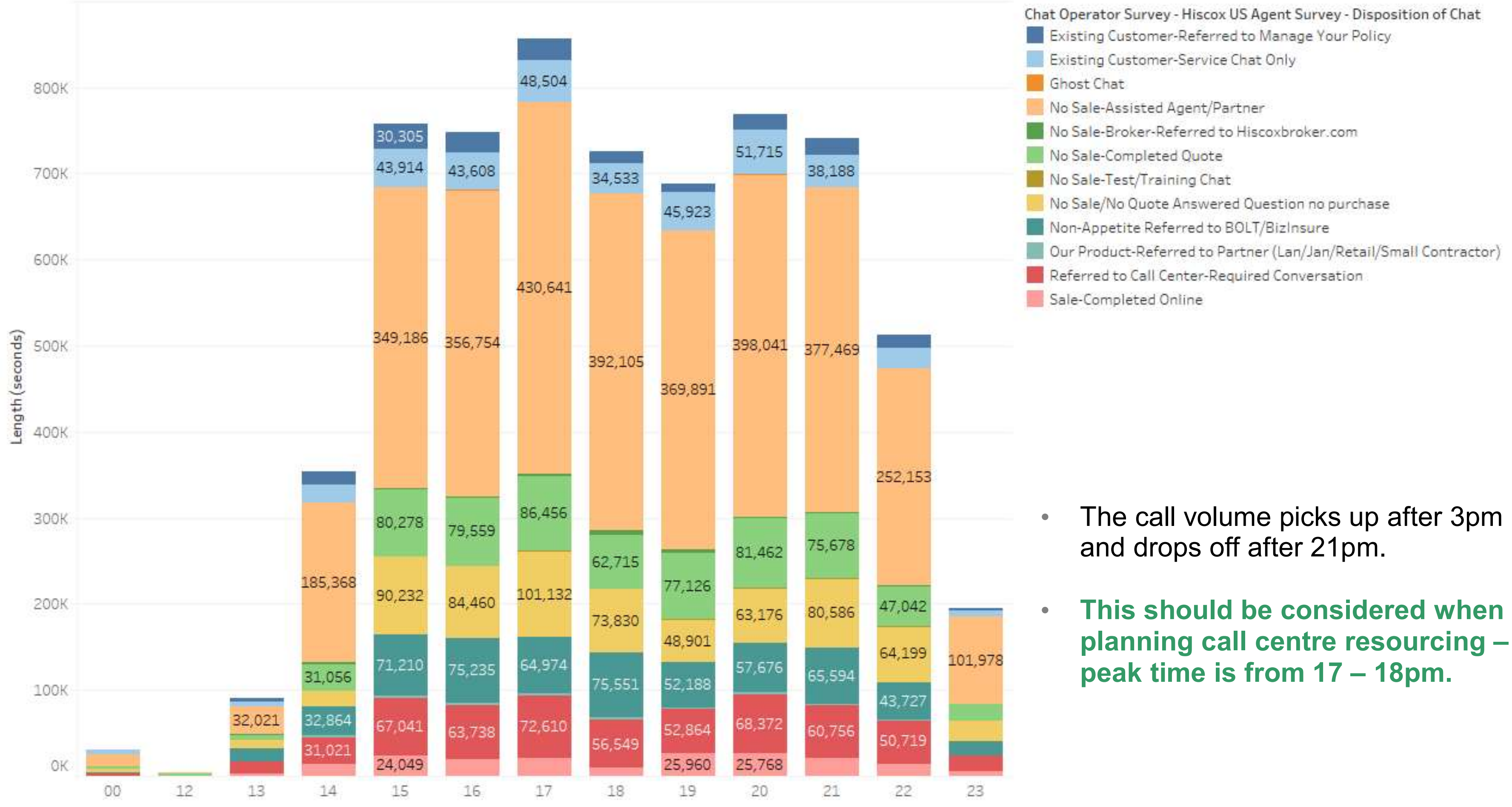


- Call volumes peak in March and then decline into the summer months.
- More call history over time is required to establish if the drop in February is incidental or a recurring phenomenon.

- **Call volumes should be taken into account when planning call centre staffing – the data suggests that more resources are required in January and from March to May while fewer may be needed during the summer months.**

Data exploration – exploring call volume by time

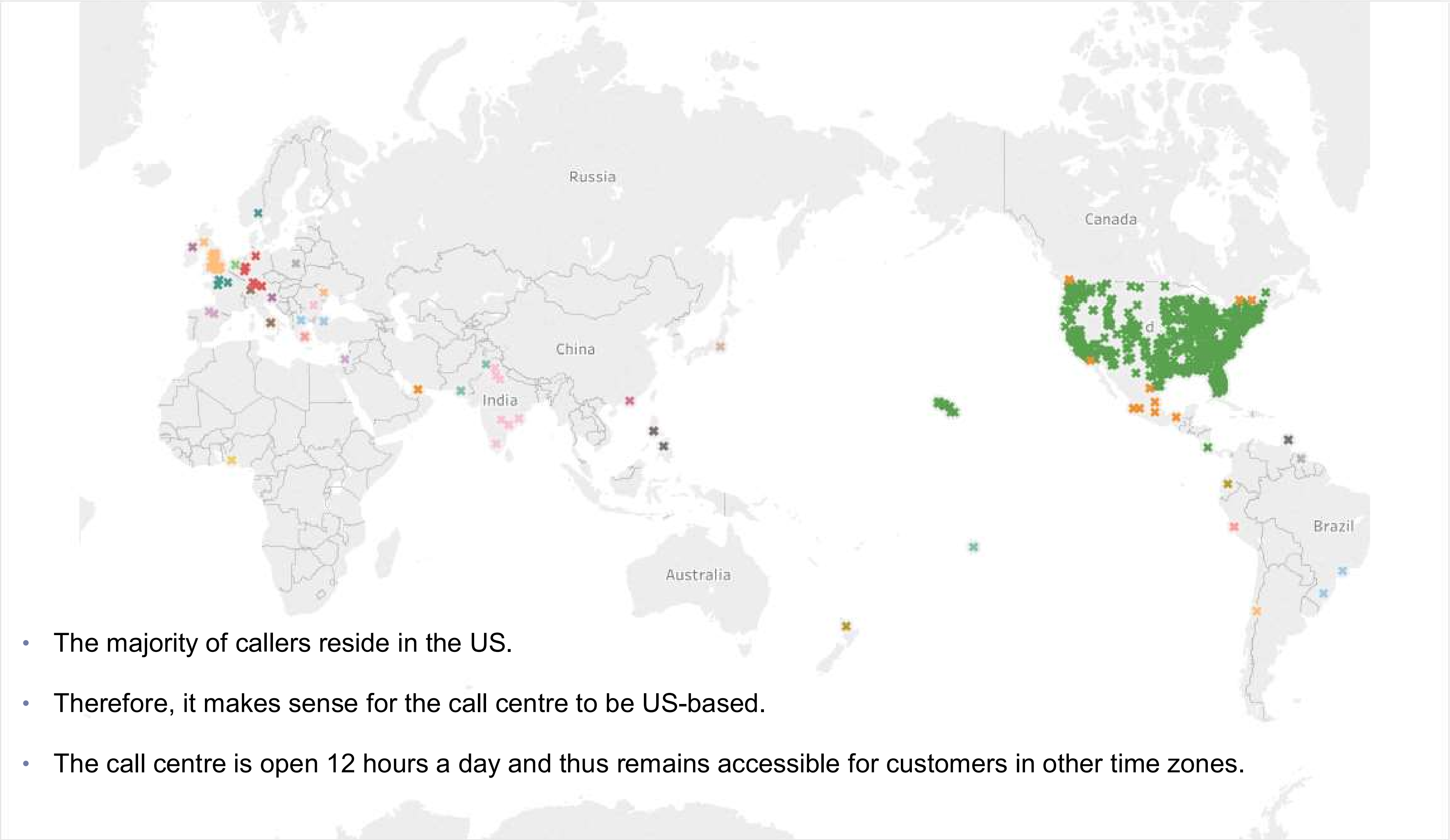
Cumulative chat length by hour of day



- The call volume picks up after 3pm and drops off after 21pm.
- **This should be considered when planning call centre resourcing – peak time is from 17 – 18pm.**

Data exploration – exploring caller countries

Caller countries



- The majority of callers reside in the US.
- Therefore, it makes sense for the call centre to be US-based.
- The call centre is open 12 hours a day and thus remains accessible for customers in other time zones.

Pre-processing

- Make every word lower case
- Tokenization

  Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens.

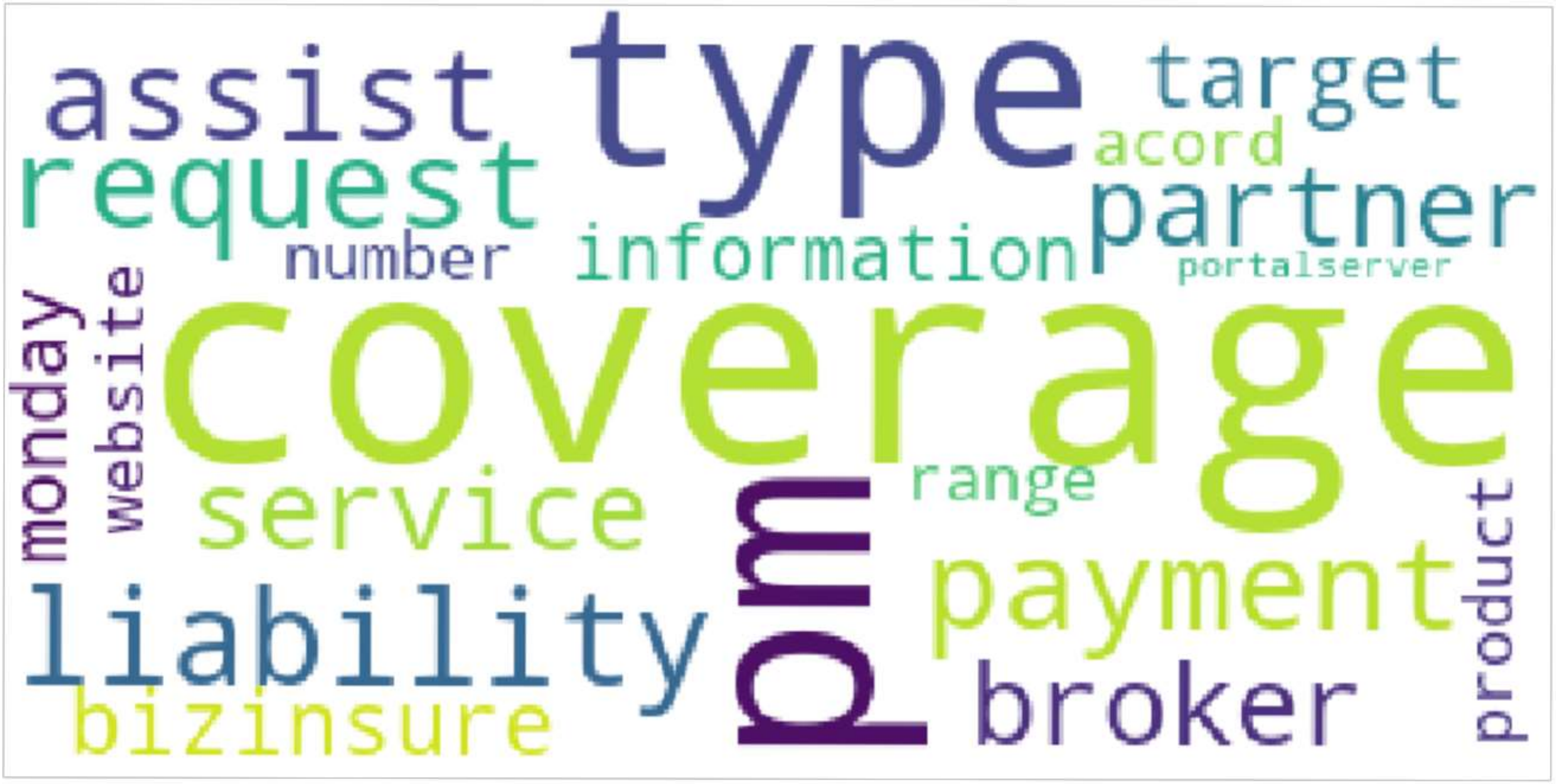
- Lemmatization

  Lemmatizaton allows the computer to group together words with the same inflected meaning but not necessarily the same lexical stem, e.g. “good”, “better”, “best”.

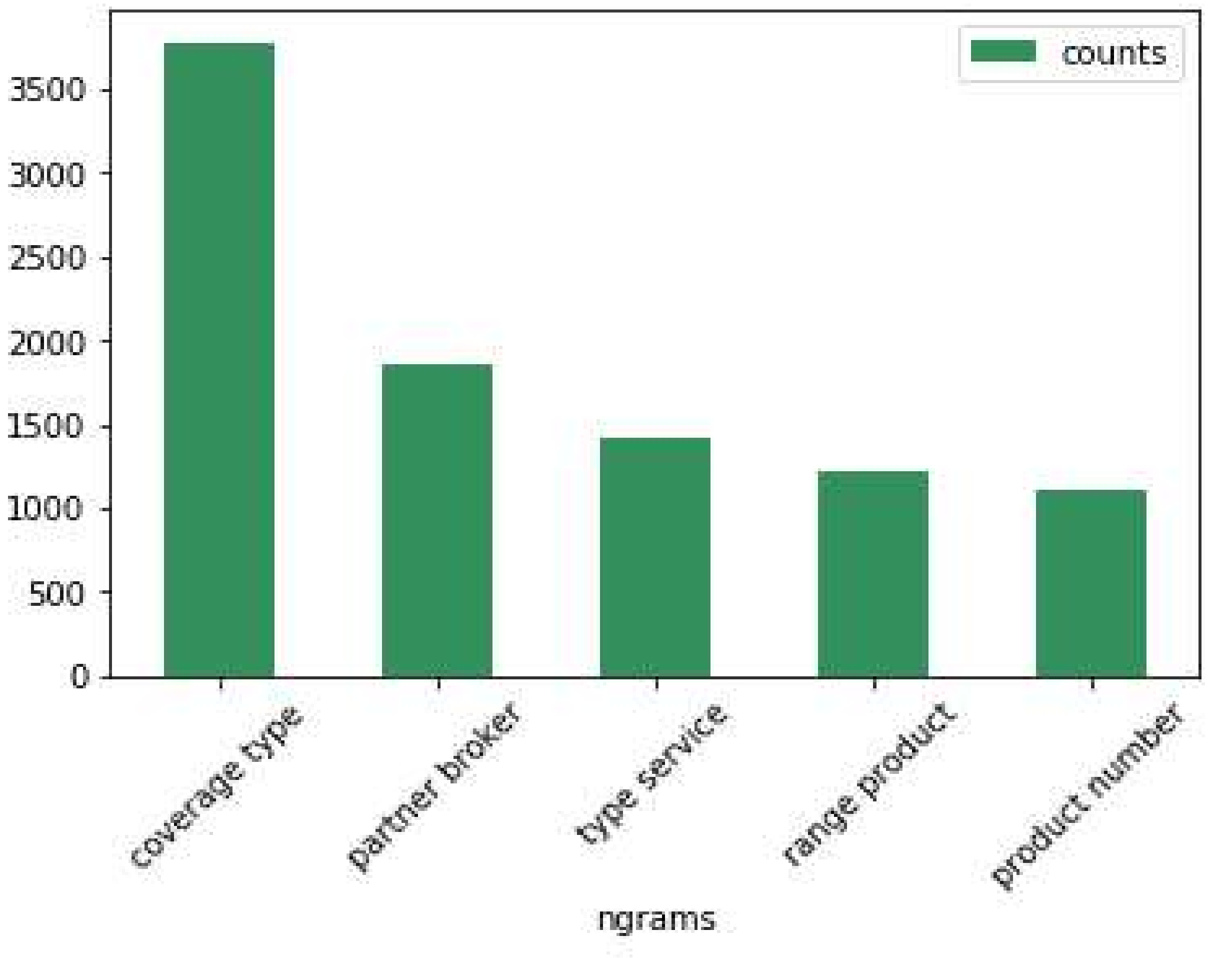
- Extend standard stopword list with first names occurring in transcript.

Most common words

- Filtered out nouns using nltk's position tagging - focusing on nouns generated the most meaningful results.
- The most common word in the transcripts is “coverage”.



Most common bi-grams



- Used nltk to extract bi-grams (i.e. pairs of words) from transcript.

- The most common bi-gram from the transcripts is “coverage type”.

- **Coverage appears to be an important word in the transcripts and warrants further investigation.**

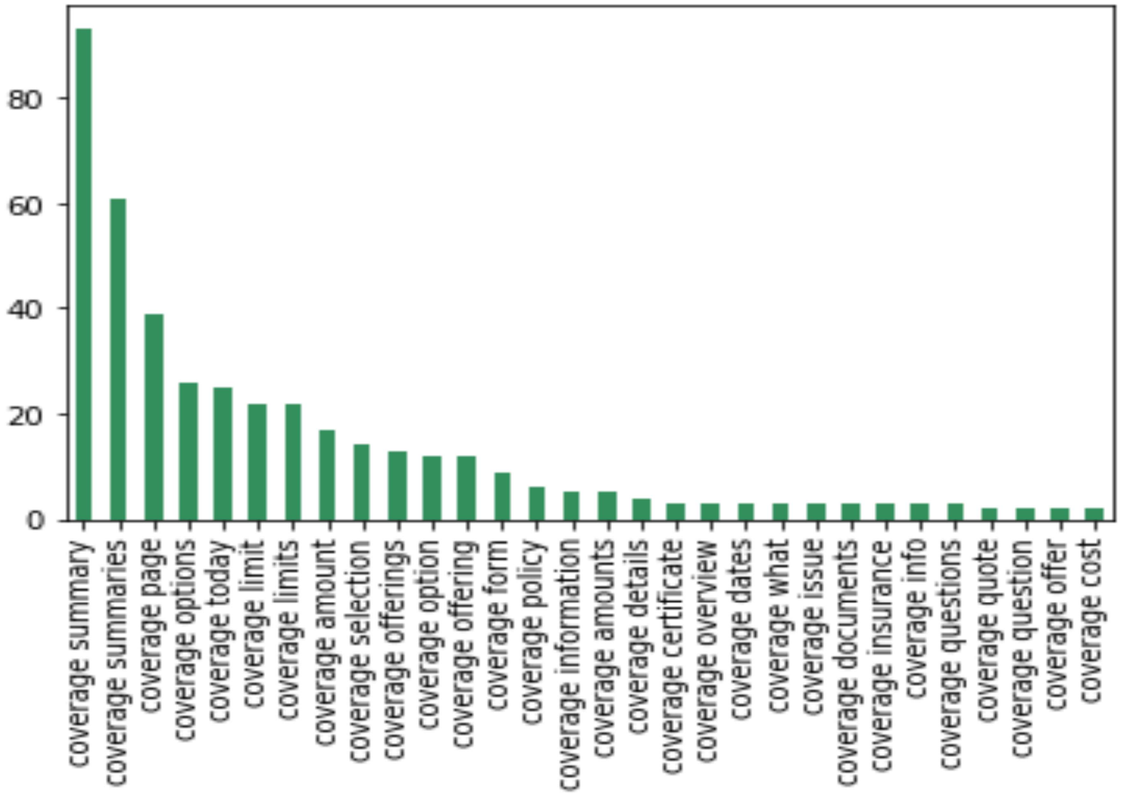
Looking at context: coverage

- Used spacy’s Matcher to extract the noun “coverage” and the subsequent noun.

- 112 matched sequences, most frequent matched sequence are versions of “coverage summary”.

Suggestion:

Liaise with website content manager to make sure sufficient information on coverage is included in the relevant sections.



Looking at context: “unable to offer coverage”

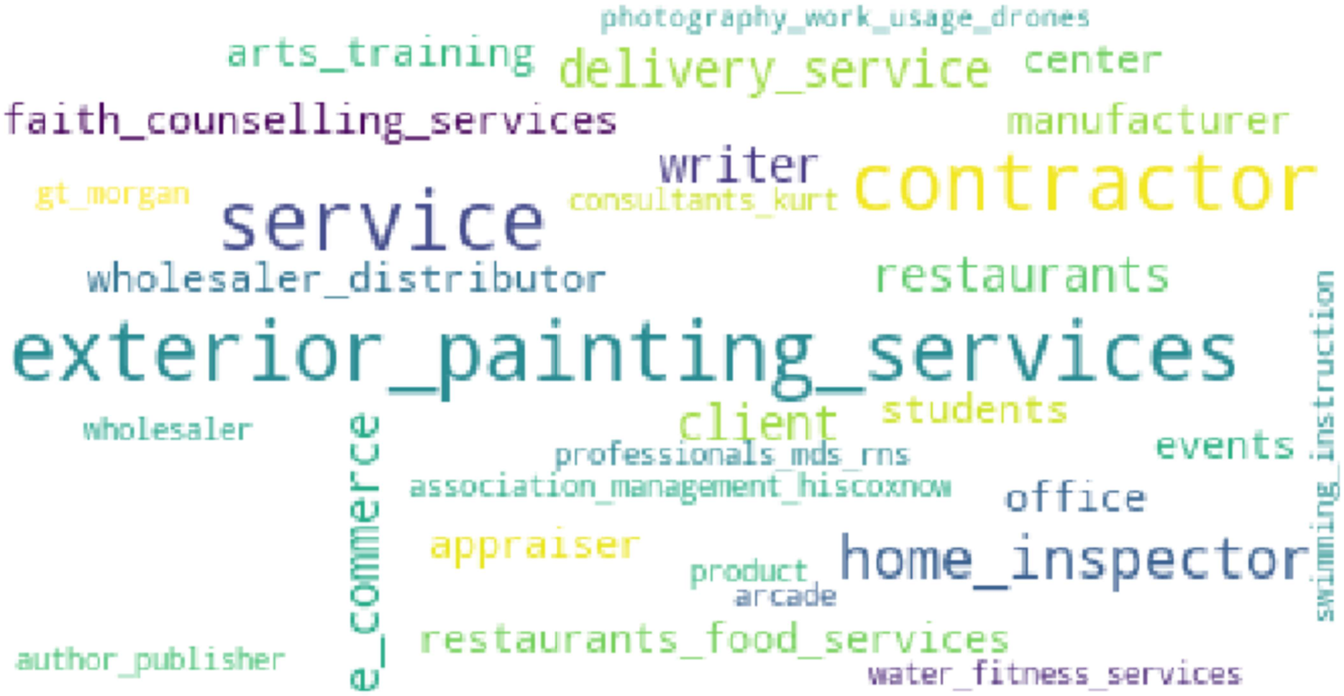
- Used spacy’s PhraseMatcher to extract phrase “unable to offer coverage” and subsequent ten words in order to find out what firm XYZ is unable to offer coverage for.

- The phrase “unable to offer coverage” is matched 2280 times.

- Within the matched phrases, exterior painting services stand out, followed by creative activities (arts training, photography work, writer, author) and restaurants.

  Suggestion:

  Liaise with Head of Product to check if it would be feasible to offer coverage for some of the below.



Most common call topics

No	Topic
1	certificate service question company document assist holder link landlord application
2	information design insured site purchase equipment compensation type owner employer
3	man liability coverage number payment option hour property documentation online

- Used gensim for Latent Dirichlet Allocation (LDA) on all transcripts to extract three topics.
- Filtered out chat disposition “Sale – Completed Online” to determine the key topics in this category.

Comment:

Within this subset of transcripts, callers seem concerned with the documentation, information on compensation and liability coverage. This is useful to know and should be brought to the attention of the website content team to ensure the documents and information are easily accessible.

Classifying chat dispositions using transcripts

- Labels: chat dispositions, data: transcripts
- Summarise the chat dispositions into 'other', 'referrals' and 'sale/quote' as some of the chat categories have very weak support.
- Train/test split: 0.67/0.33.
- Used sklearn's CountVectorizer to convert the transcripts to a matrix of token counts.
- Used sklearn's Linear Support Vector Classifier.
- Train classifier on training set, evaluate performance on test set.

Classifying chat dispositions using transcripts : assessing performance

	Precision	Recall	F1-score	Support
Other	0.93	0.96	0.95	3116
Referral	0.86	0.85	0.85	522
Sale/quote	0.54	0.35	0.42	237
Micro avg	0.91	0.91	0.91	3875
Macro avg	0.78	0.74	0.74	3875
Weighted avg	0.90	0.90	0.90	3875

- As the support column shows, the dataset is unbalanced. The classifier requires a certain level of support to predict well.
- Support for sale/quote in the test test is weak, which explains the poor performance of the model in this category. However, performance in the other two categories with better support is acceptable.

Suggestion:

Get more data and re-run.

Conclusions & recommendations

- The average call time spent on the different call dispositions seems intuitively right.
- Agent Rebecca appears to be the more productive of the two agents, provided both work the same amount of hours.
- When planning call centre staffing, take into account the variation of call volume by hour of day and month of year. In a day, the call volume is highest between 15 and 21pm, reaching a peak between 17 and 18pm. While it is more difficult to confirm the existence of seasonality for certain due to call transcripts from January to August only, an increase in call volumes in late spring can be observed, which is decreasing towards the summer months.
- “Coverage” is the most frequent word in the text body, and “coverage summary” is the most frequent bi-gram. Given the importance of the term, it is recommended to liaise with the website content team to review the information provided on policyholders' coverage and possibly enhance this.
- In terms of product development, it is important to know what coverage firm XYZ is not able to offer. Exterior painting services stand out, followed by creative activities (arts training, photography work, writer, author) and restaurants. It is recommended to liaise with product development to review the coverage offered for certain business types – it may be feasible to extent coverage in certain areas and thus write more business.
- Topic extraction is an effective way to extract the key subjects from the transcripts, especially when filtered according to chat disposition.
- Text classification can be used for automated labelling and could form part of an automated call transcript analysis within firm XYZ. Note that more data is required for reliable model performance.

Appendix I

Latent Dirichlet allocation (LDA) is a probabilistic topic model that assumes documents are a mixture of topics and that each word in the document is attributable to the document's topics. LDA defines each topic as a bag of words, and you have to label the topics as you deem fit. There are 2 benefits from LDA defining topics on a word-level: 1) We can infer the content spread of each sentence by a word count: **Sentence 1**: 100% Topic F **Sentence 2**: 100% Topic P **Sentence 3**: 33% Topic P and 67% Topic F 2) We can derive the proportions that each word constitutes in given topics. For example, Topic F might comprise words in the following proportions: 40% eat, 40% fish, 20% vegetables, … LDA achieves the above results in 3 steps. Step 1 **You tell the algorithm how many topics you think there are.** Step 2 **The algorithm will assign every word to a temporary topic**. Step 3 (iterative) **The algorithm will check and update topic assignments**, looping through each word in every document. For each word, its topic assignment is updated based on two criteria: How prevalent is that word across topics? How prevalent are topics in the document?

Source: algobeans.com

Appendix II

The key differences between LinearSVC and SVC are as follows:

By default scaling, LinearSVC minimizes the squared hinge loss while SVC minimizes the regular hinge loss. It is possible to manually define a 'hinge' string for loss parameter in LinearSVC.

LinearSVC uses the One-vs-All (also known as One-vs-Rest) multiclass reduction while SVC uses the One-vs-One multiclass reduction. It is also noted here. Also, for multi-class classification problem SVC fits N * (N - 1) / 2 models where N is the amount of classes. LinearSVC, by contrast, simply fits N models. If the classification problem is binary, then only one model is fit in both scenarios. multi_class and decision_function_shape parameters have nothing in common. The second one is an aggregator that transforms the results of the decision function in a convenient shape of (n_features, n_samples). multi_class is an algorithmic approach to establish a solution.

The underlying estimators for LinearSVC are liblinear, that do in fact penalize the intercept. SVC uses libsvm estimators that do not. liblinear estimators are optimized for a linear (special) case and thus converge faster on big amounts of data than libsvm. That is why LinearSVC takes less time to solve the problem.

Source: https://stackoverflow.com/questions/45384185/what-is-the-difference-between-linearsvc-and-svckernel-linear