Graduate School of Management
St. Petersburg University

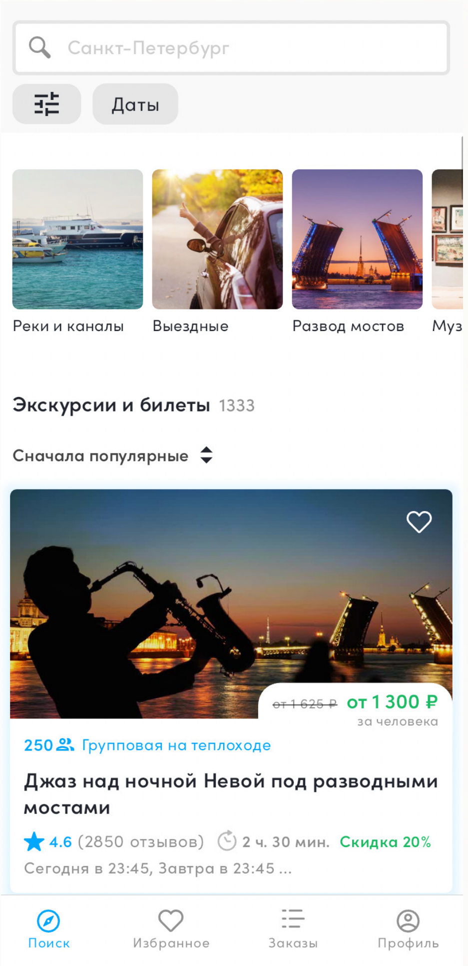
SPUTNIK

Increasing responsiveness to customer feedback by applying opinion mining in travel tech

Elizaveta Pavlova, MIBA Master Program
Academic Advisor – Vladimir Gorovoy

SPUTNIK is the largest worldwide tour booking service in Russian.

Being a travel-tech company, it allows tourists to search guided tours, sightseeing options and personal guides, and book activities online, in advance or on the go.



Given

Data management tools:

- Data warehouse (DWH)
- BI tool – MetaBase

Stakeholders

Supply, Product, Marketing, Content, etc.

User's pains

- untapped source of insights
- unstructured, freeform nature of reviews
- different components of customer experience
- mistakes in reviews
- common themes
- lack of aggregation

Research Questions

RQ1. How can the textual customer reviews be most efficiently translated into business insights?

RQ2. Which measures can be applied to fine-tune the review analysis algorithm?

RQ3. By how much can the review analytics tool improve the reaction to customer feedback?

Research Goal

Devise integrated solution for Feedback Analytics for Sputnik8 business and data context

Context

What constitutes Tourist Experience in travel industry?

Can customer repurchase be predicted from review text?

How can customer reviews be better broken down into product aspects?

Literature

Existing publications on theoretical grounds

Research

Synthesis of conducted research and carried out overview

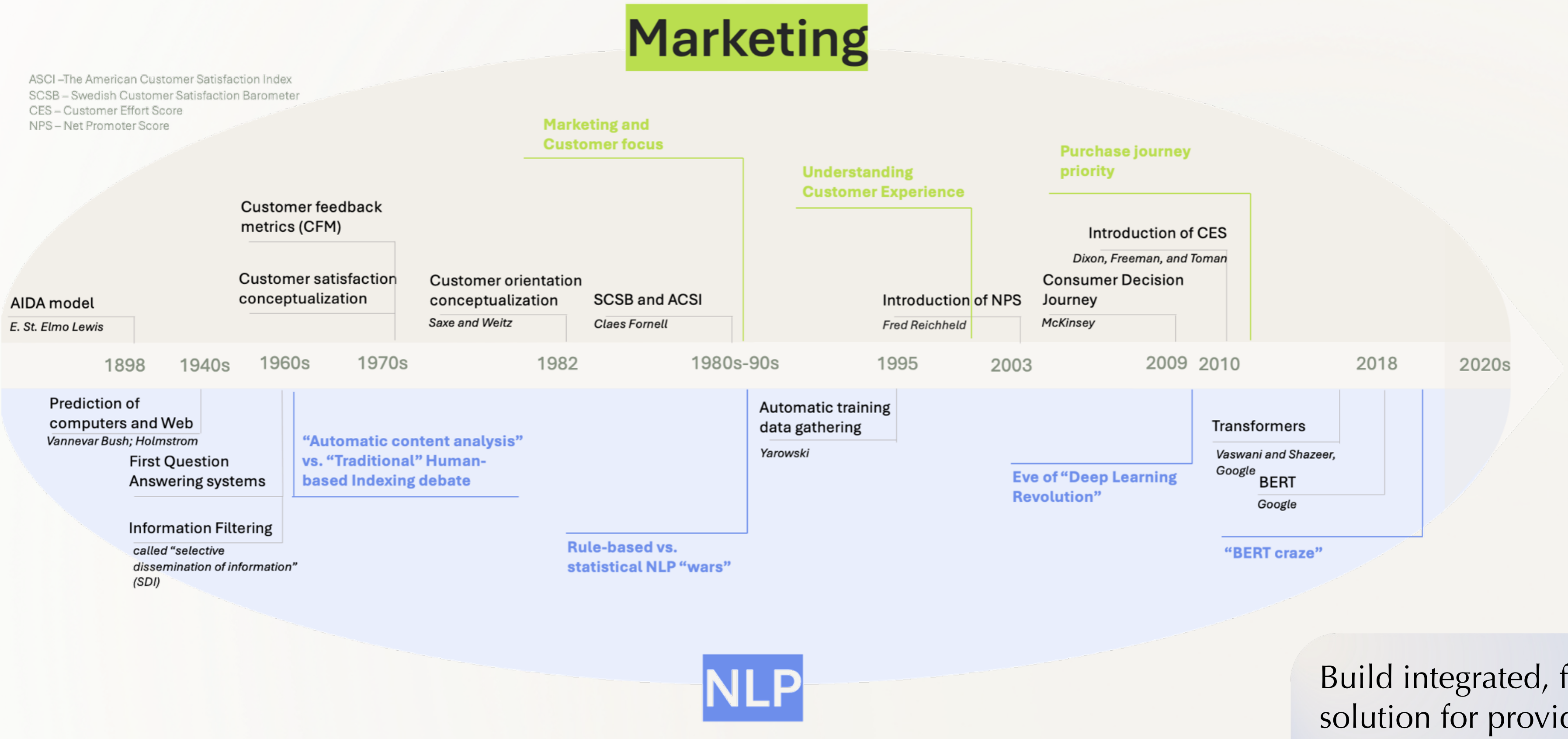
Theory

Quantification of Marketing Impact and Customer Feedback Metrics

What is state of the art in text classification?

Advantages and drawbacks of NLP classification approaches

Literature Review & Research Gap



Research Gap:

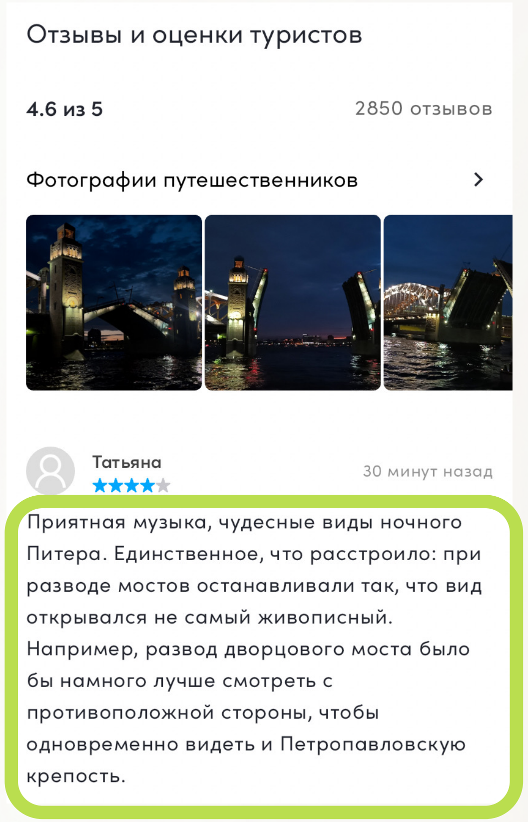
1. Understudied review analytics for **excursions segment** of travel
2. Lack of **computationally efficient** review analytics solutions in Russian applicable to **real business tasks**
3. Lack of **customer feedback DataViz**

Build integrated, fast and **computationally efficient** solution for providing timely analytics of reviews in Russian, which:

- categorizes reviews by aspects
- predicts repurchase

362.7K unique rows

acquired directly from corporate database



Raw Review

Text Preprocessing
emoji
special & junk characters
spelling inconsistencies

Text Embedding

Repurchase Classification

Sentence Segmenting

Text Embedding

Aspect Classification

	Repurchase Classification	Aspect Classification	Polarity Classification
ML task type:	Binary Classification	Multi-Class Classification	Binary / Multi-Class Classification
Question to answer:	*Will customer make a repeat purchase after leaving the review?*	*Which component(s) of service does the review assess?*	*Is emotional tone of message positive, neutral, or negative?*
Training data:	Juxtaposed historical order data	Manual Labelling	Review Rating Score: from 1 to 5 stars

Model chosen for retraining:
cointegrated/rubert-tiny2

- trained exclusively for Russian language
- provides tokenizer with supported sequence lengths up to 2048
- fast-performing

Repurchase Classification

13 months

Review Posted — Repurchase — Not repurchase

Why 13 months?

- Travel industry seasonality
- Customer's vacation seasonality

13 months back

all reviews — only reviews with further repurchase — **Training Set Formation**

Aspect Classification

Manual tagging with simultaneous class definition
Number of resulting classes **unknown**

Polarity Classification

Multi-Class Classification: 1★ 2★ 3★ 4★ 5★

Binary Classification: **negative** **positive**

GitHub Project Repository

Evaluation metric:
F-score (macro-averaged)

Repurchase Classification

0.90 Final

0.61

0.52

Aspect Classification

0.94 Final

0.73

0.57

0.34

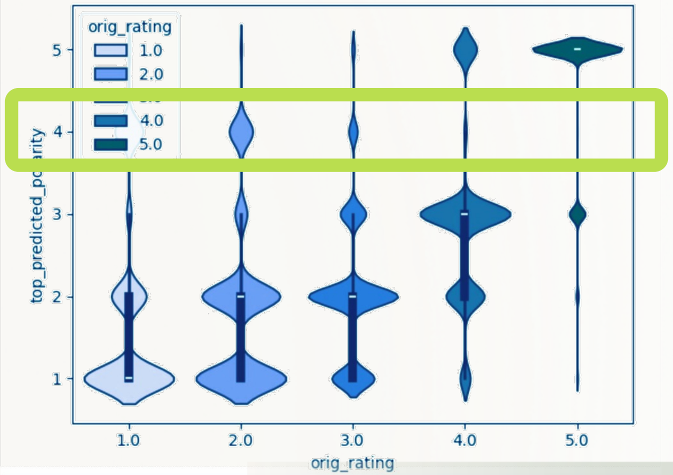
Polarity Classification

0.87 Final

Binary

0.34* Final

Multi-Class



* Explicit customer ratings **cannot be** relied upon to express sentiment

Review Aspects & Issues
For Aspect Classification

	sputnik	guide	organization	tour content	facilities	beyond control	subjective
Integral Issues Highlight growth points; Can be influenced; To be controlled	hazard customer service	rudeness & bigotry coordination guide's proficiency theme divergence fact misstatement	time shortage price lift haste	offer disparity no substance overpriced humble route			
Situational Issues Can be single-time; If repeated, indicate systemic problem; Some control	refund cancellation		delay program change stray		vehicle dirty windows conditioning	other tourists weather	weather | advice
Perceptual Issues Tourist's perception; Close to no control		guide's attitude shallow narration bland delivery			food		unmet expectations appraisal unrelated

Justifies sentence segmenting

Single review tends to cover a number of issues, belonging to different aspects

prediction score

Тема до конца не раскрыта **OFFER DISPARITY** 98% Все очень поверхностно и быстро, информации было меньше чем в открытых источниках в интернете **SHALLOW NARRATION** 90% Гостиницу Ленинградская и высотку на Красных воротах вообще мимоходом проехали **UNMET EXPECTATIONS** 50% Гид больше пытался шутить неудачно и не смешно чем раскрывать тему экскурсии **GUIDE'S PROFICIENCY** 55%

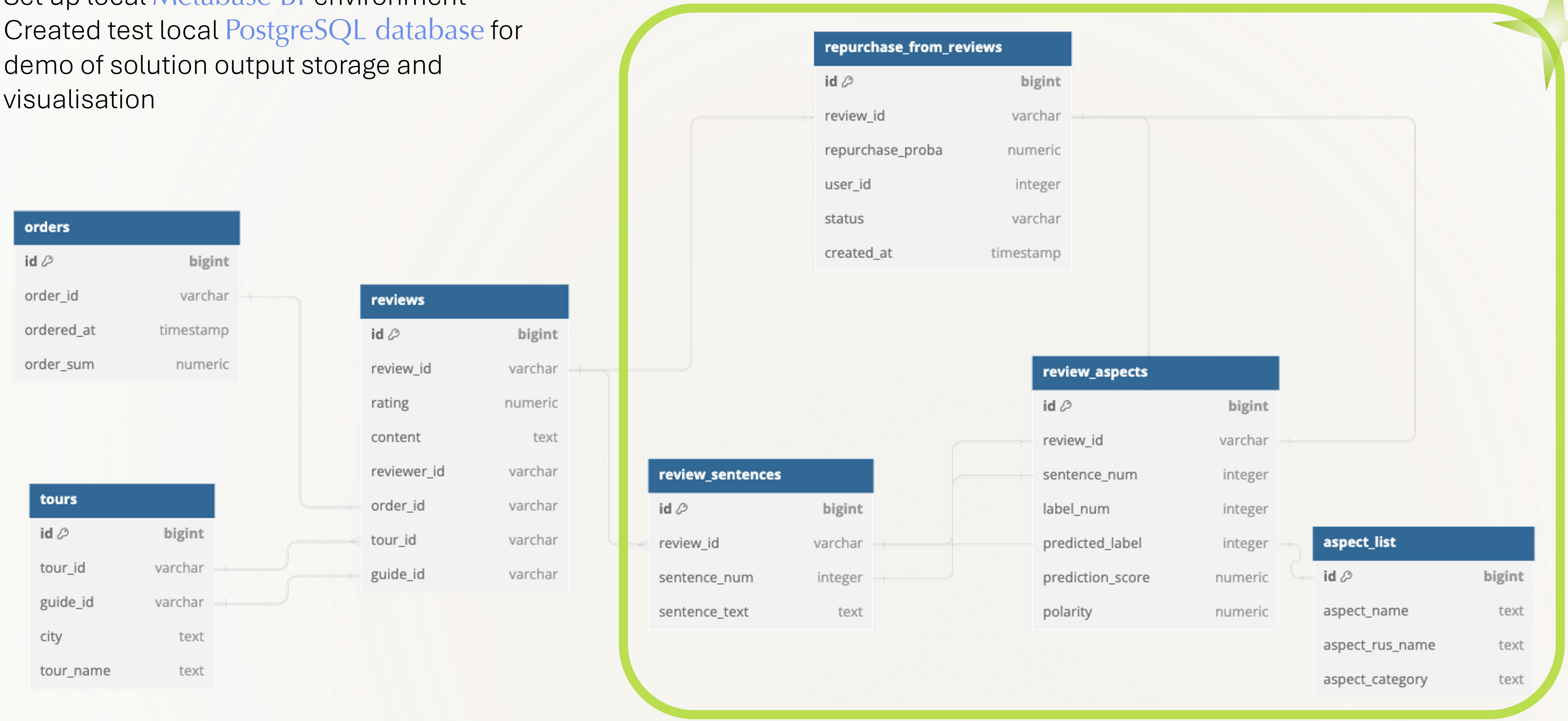
predicted aspect

Были проблемы с подтверждением оплаты за экскурсию с организатором, пришлось обращаться в Спутник **CUSTOMER SERVICE** 99% Обед в посёлке, с дегустацией, к сожалению без выдачи квитанций и чеков **FOOD** 54% Экскурсовод отрабатывает свои деньги на всё 100 **COORDINATION** 64%

Хотелось по дорогое как при других экскурсиях получить больше информации о городе, о достопримечательностях **NO SUBSTANCE** 87% Так как гида на фонтанах не было, быстро почти молча пробежали по части территории, личного времени на фото и прогулку выделено мало **HASTE** 96% Нужно хотя бы 40-50 мин **DELAY** 99%

Отвратительно выстроен тайминг **COORDINATION** 93% На косе совершенно не хватило времени ни к морю сходить, ни поесть **TIME SHORTAGE** 99% В Зеленоградске тоже голопом проскакали и уехали **DELAY** 88% Сами места прекрасные, гид рассказывает хорошо, но тайминг испортил всё впечатление **APPRAISAL** 95%

Set up local Metabase BI environment
Created test local PostgreSQL database for demo of solution output storage and visualisation



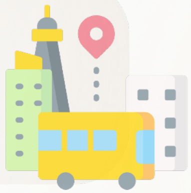
Dashboard Design Drivers

- What questions would dashboard users seek to answer?
- How to present these answers in most digestible, actionable form?

Dashboard Components

Entities

Excursion
Particular tour offer



Guide
The supplier of excursion

Review
The feedback on order

Dimensions

Tour
Guide
City
Time
Aspect / Aspect Category

Measures

repurchase probability
aspect frequency
calculated measures

After which tours customers will more likely buy again?

Repurchase Likelihood by Tour

This question is written in SQL. Start Date City

TOUR_ID	City	Avg Repurchase Likelihood	Avg Rating
RPQO81338511263041	Париж	99%	3.67
JSHK13908176818653	Вена	94%	4.67
IHAB37111947596201	Санкт-Петербург	90%	4.56
IAKG72251639871072	Воронеж	83%	5
WBCK02087150800334	Санкт-Петербург	80%	4.6
LZIP62665654460394	Таллин	77%	5
CCUR53744232408374	Прага	76%	4.82
TMPA33680683324335	Санкт-Петербург	76%	4.83
QUBI82336890832008	Москва	75%	4.5
TXFT70842137804044	Римини	74%	3.58

[Product + Supply] Working on customer journey transformation, leading user to buy more preferable tour first

Dashboard Demo & Use Cases

Aspects by Tour

This question is written in SQL. | Top N | Since | Tour ID

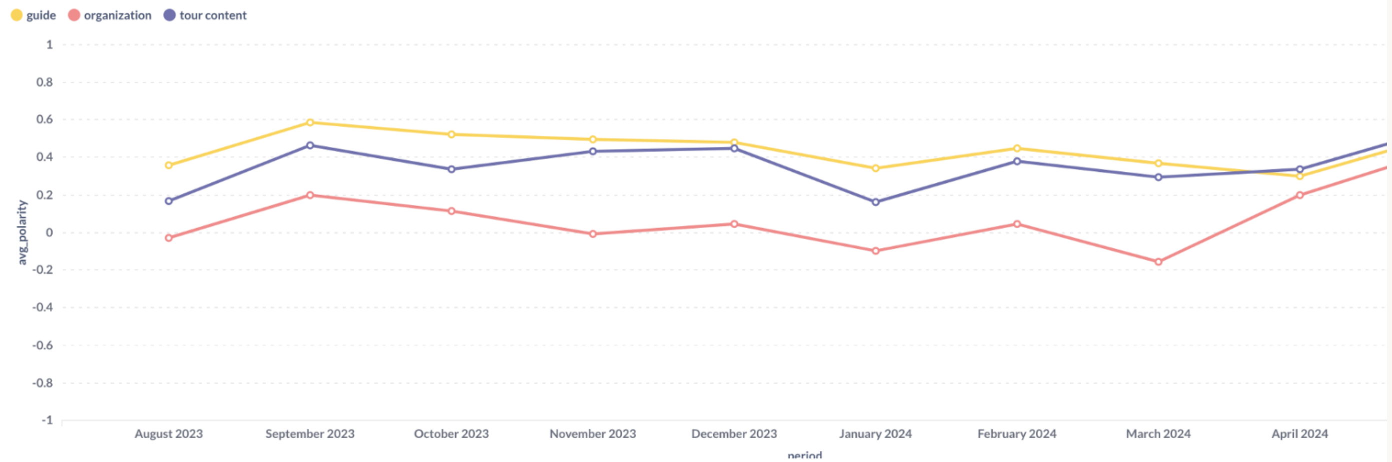
City	Tour ID	Rank	Aspect	% of Reviews	# Mentions
Зеленоградск	AAAZ03251383027785	1	мастерство гида	72%	18
Зеленоградск	AAAZ03251383027785	2	мнение	64%	16
Зеленоградск	AAAZ03251383027785	3	отношение гида	52%	13
Зеленоградск	AAAZ03251383027785	4	содержательность	48%	12
Зеленоградск	AAAZ03251383027785	5	поверхностный рассказ	44%	11
Сочи	AAEG96671245223491	1	мастерство гида	85%	11
Сочи	AAEG96671245223491	2	мнение	85%	11
Сочи	AAEG96671245223491	3	неоправданные ожидания	77%	10
Сочи	AAEG96671245223491	4	отношение гида	77%	10
Сочи	AAEG96671245223491	5	содержательность	69%	9

What characterizes the tour,
how customers tend to describe it?

[Supply/Content] Use most common aspects by Tour for selecting a tour for further promotion
[Supply] Devising new Tour categories and groupings
[Product] Using aspects as tags for ranking and recommendation

Dynamics over Time - Tour Polarity by Aspect

This question is written in SQL. | Since: August 1, 2023 | Tour ID: HXHN30298898850215 | Agr Date: month | Aspect Category: 3 selections



(For poorly performing Excursion)
What customers find most unpleasant in tour, and how can we fix it?

[Supply] Find most problematic aspect
[Supply] Track it through time, to identify when the tour got worse, and approach guide for improvement

Business Effect: Monetary

Revenue about to be Lost - Repurchase by Tour

This question is written in SQL.

Since: August 1, 2023 ×

City

City	Tour ID	Potentially Losing
Санкт-Петербург	OTGP89013311281478	642,000
Санкт-Петербург	KGKY47044651546854	331,000
Санкт-Петербург	FGWM21535456721591	328,000
Дубай	ZRVG02105368176640	309,000
Махачкала	CCGQ93363933725246	241,000
Москва	RTHY06125045820290	227,000
Санкт-Петербург	VISA49523102058725	200,000
Санкт-Петербург	RAIZ37287867195613	171,000

Assessing foregone earnings

How much revenue could the customers bring the company if they were satisfied with their previous tour experience?

$$Revenue_{lost} = \sum_{i=1}^{n} Cost_{order\ i} * Repurchase\ prob_{review\ i}$$

Business Effect: Responsiveness

Modeling a case for 1 popular excursion that starts to gradually decline in rating

- *By when the problem will be noticed on rating alone?*
- *How long will it take Supply manager to read through negative feedback to identify the problem?*

Cost of delay

1. Postponing problem solving stage
2. Employee's time spent inefficiently
3. Potentially losing customers with negative experiences and lifetime revenue from them

It takes 13 ms - 150 ms to process a visualization

		Week 1	Week 2	Week 3
Weekly # of Reviews	120			
Avg Reading Speed for sentence	3.75*			
Weekly Rating (Week 0)	4.72	4.6	4.5	4.2
Avg # of Sentences				
	4.72	114	108	96
4.95	3	3	6	12
6.68	2	2	4	8
6.98	1	1	2	4
Sum # of Sentences		35	70	141
seconds reading		131.96	263.93	527.85
cumulative seconds reading			395.89	923.74
minutes reading		**2.20**	**6.60**	**15.40**

** - Tuscher, Michaela, and Johanna Schmidt.*
Processing Speed and Comprehensibility of Visualizations and Texts. 2022.

Have built integrated, fast and computationally efficient solution for providing timely analytics of reviews in Russian, which:

- categorizes reviews by aspects
- predicts repurchase

- Defined product aspects based on customer reviews
- Trained 3 different-purpose classification models (**F-score ~0.9**)
- Outlined required database design
- Built dashboard prototype, highlighting use cases and actionability
- Estimated time to be saved by implementing the proposed solution

RQ1. How can the textual customer reviews be most efficiently translated into business insights?

Implementing probabilistic predictions (repurchase); categorizing unstructured data (aspects)

RQ2. Which measures can be applied to fine-tune the review analysis algorithm?

Adding training data; improving data labelling; retraining in stages; adjusting hyperparameters

RQ3. By how much can the review analytics tool improve the reaction to customer feedback?

Reducing Review Analytics stage from **~15 mins** to **Dashboard load time**

Output

BUY

3 Trained Models

Data Design

Dashboard Prototype

Thank you, that is all.
Questions?

Increasing responsiveness to customer feedback by applying opinion mining in travel tech

Elizaveta Pavlova, MIBA Master Program
Academic Advisor – Vladimir Gorovoy

Reviewer's Question 1

Do the repurchase classifier and polarity classifier generate dissimilar outputs? One can assume that review sentiment and reviewer's intention to make repeat purchases correlate. To what extent do they correlate, and why was it necessary to make them into two separate models?

Provide differing predictions

polarity:	negative	positive	All
repurchase:			
won't repurchase	95 184	147 052	242 236
will repurchase	47 798	71 554	119 352
All	142 982	218 606	361 588

polarity:	negative	positive	All
repurchase:			
won't repurchase	26.3%	40.7%	67.0%
will repurchase	13.2%	19.8%	33.0%
All	39.5%	60.5%	100.0%

Jaccard Similarity coefficient score: **0.269**

Cosine Similarity: **0.443**

Conceptually different

Customer can make repeat purchase after an unsatisfactory experience, or not buy again with us even when they are satisfied

Reviewer's Question 2

Why was the aspect model ultimately realized as a classification one? Considering there were no ground truth labels on data, and the number of classes was unknown, would it not be more suitable to cluster the reviews rather than classify them?

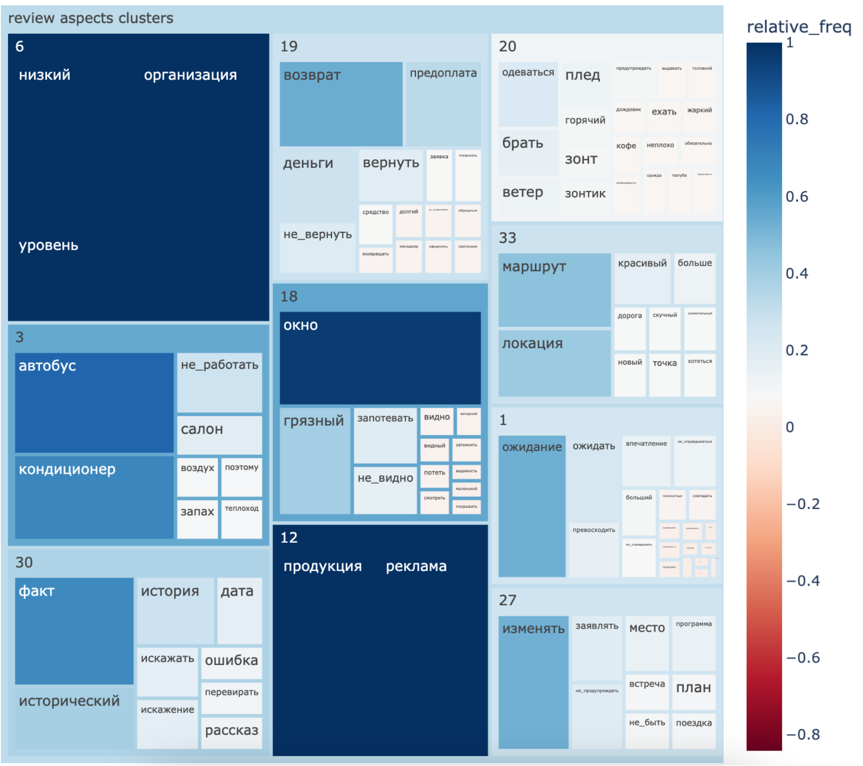
Justification

- number of classes was unknown, but was expected to be fixed
- rareness of text clustering
- more adjustable training process

Clustering was used as supplement

KMeans clustering model with 36 clusters, which:

- **facilitated EDA**: finding word frequencies and most common words within clusters
- **supported manual labelling decision**: matching Clustering and Classifier predictions were used as training data with more certainty



Common Word Treemap by Cluster

Reviewer's Question 3

Can the models be applied for analyzing other textual data, such as clients' inquiries to customer support? What changes would need to be introduced to make a solution versatile to process all types of customer feedback?

Retraining would be required

UGC for predictions:

- Inquiries to customer service
- Chat messages
- App Store reviews
- Social Media Comments

Should take into account

- Each type of UGC has their own aspects (App Store Reviews – product ones, Chat Messages – mostly contain questions)
- Channel-specific idiosyncrasies (syntax, lexis)

Further Development Directions

- **Keyword search** (through stemming and frequencies)
- **Helpfulness** ranking
- Review **Summary**
- Hazard **Alerts**
- etc.

model	CPU	GPU	Size	Mean S	Mean S+W
cointegrated/rubert-tiny2	4.9	2.7	112	0.689	0.631
DeepPavlov/rubert-base-cased-sentence	100.6	7.6	678	0.656	0.594
sberbank-ai/sbert_large_nlu_ru	348.6	14.2	1590	0.654	0.599

The comparison of model speed, size, and performance for the three considered pretrained embedders (Dale, 2022)