

INSIGHT EXTRACTION

reMarkable



Alena, Arshak, Daniel, Krissy, Oliver, Rana

TABLE OF CONTENTS

01

Business Understanding

Company background
and situation assessment

03

Modeling

Deriving insight from data

05

Dashboard

Tracking and monitoring
KPIs

02

Data

Data collection, cleaning
& descriptions

04

Analysis Outcome

Insights and
recommendation

06

Next Step

Connections with reMarkable
VP of Marketing

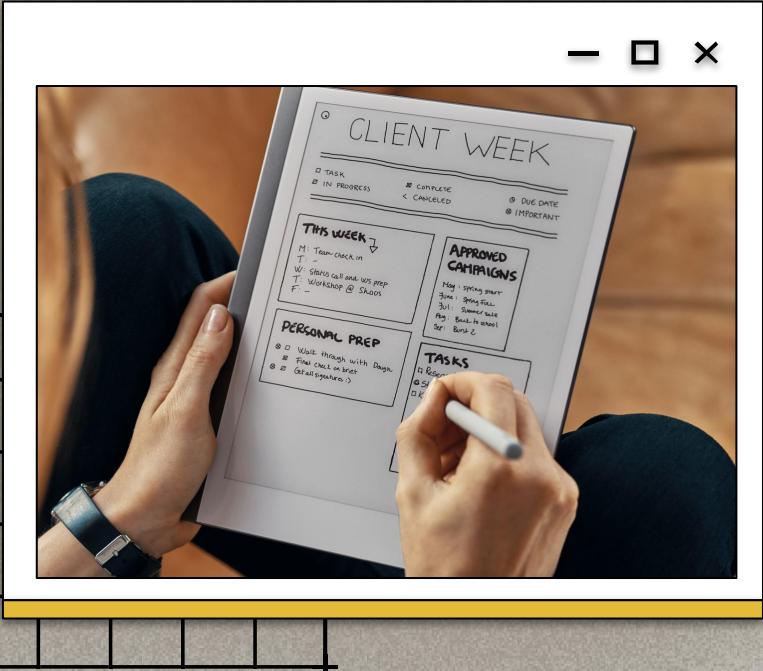
01

BUSINESS OBJECTIVES

The Focus for this Project



Background



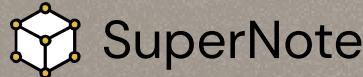
Norwegian-Based



Electronic Ink-Based Tablet

Background

Main Competitors:



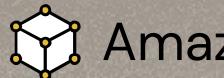
SuperNote



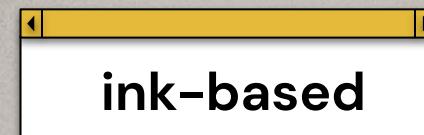
Boox



Apple iPad



Amazon Kindle





Main Objectives

PRODUCT DESIGN

identify specific and actionable customer needs

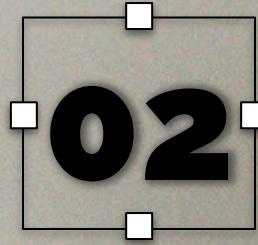
MARKETING

effectively target the audience interested in our products



SUCCESS CRITERIA

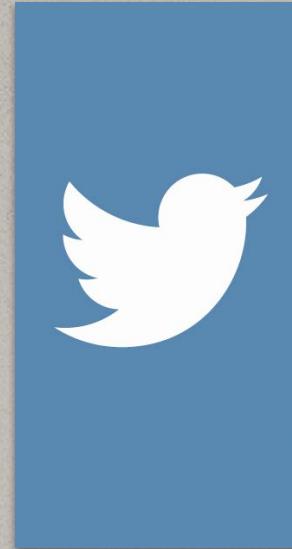
increase product sales by 20% compared to reMarkable 2 and achieve higher customer satisfaction.



DATA

Data Collection and Data Cleaning

Data Collection





Twitter API



10,000,000 Tweets



Back to 2006
vs 7 days

Twitter

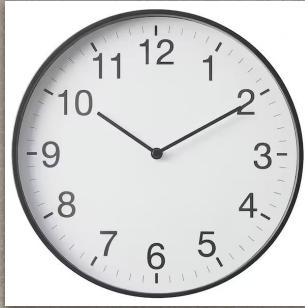


Twitter API

Academic Research product track

Advance your research objectives with public data on nearly any topic.

Libraries



Data Collection Steps



Craft an advanced query

```
1 # specify the query
2 query = 'remarkable2 lang:en -is:retweet'
3
4 # specify the period
5 # this variable can be used in a loop to get tweets from different months
6 # to see trends in sentiment or in topics
7 start_time = '2010-01-20T00:00:00Z'
8 end_time = '2022-04-28T00:00:00Z'
```



Custom Pagination



Data Collection Steps



Custom Pagination

```
for page_number in range(n_pages):
    if next_token != 'null':

        # request the tweets from the API
        tweets = client.search_all_tweets(query=query,
                                          tweet_fields= tweet_fields,
                                          expansions = expansions,
                                          user_fields=user_fields,
                                          place_fields = place_fields,
                                          start_time = start_time,
                                          end_time = end_time,
                                          next_token= next_token,
                                          max_results=n_per_request)
```

request the
“book”

```
try: # if there is a next page then continue
    print("next_token:", tweets.meta['next_token'])
    next_token = tweets.meta['next_token']

# otherwise break
except: break
```

“page number”

Data Collection Steps



Respect the server

429

Too Many Requests

v1.1

Returned when a request cannot be served due to the App's rate limit or Tweet cap having been exhausted.
See Rate Limiting.



```
# respect the server, take it slow, sleep for 1 second after each request  
time.sleep(1)
```

Data Collection Steps



Tweets and Users

```
# Get users list from the includes object
users = {u["id"]: u for u in tweets.includes['users']}

# loop through each tweet
for tweet in tweets.data:

    # get the user based on author id from the tweet
    user = users[tweet.author_id]
```

Data Collection Steps



Combine the data

```
# append the data from one tweets to the dataframe
all_tweets_df = all_tweets_df.append({
    'tweet.id':tweet.id,
    'tweet.text':tweet.text,
    'tweet.attachments':tweet.attachments,
    'tweet.author_id':tweet.author_id,
    'tweet.context_annotations':tweet.context_annotations,
    'tweet.conversation_id':tweet.conversation_id,
    'tweet.created_at':tweet.created_at,
    'tweet.entities':tweet.entities,
    'tweet.geo':tweet.geo,
```

Data Collection Steps



1 line → 10,000,000 Tweets

```
] 1 # get all the tweets by query, number of tweets, start_time, end_time
2 all_tweets_df = get_tweets(query, 100000, start_time, end_time)
3

page: 1 , tweets: 0 - 100
next_token: b26v89c19zqg8o3fpytma5hsfc1q16174lipm9sw6z14t
page: 2 , tweets: 0 - 200
next_token: b26v89c19zqg8o3fpytma3froh16cblc0yxx8nukfhchp
page: 3 , tweets: 0 - 300
next_token: b26v89c19zqg8o3fpytma3frieynekyyqg0dibn50t7r1
page: 4 , tweets: 0 - 400
```

DATA CLEANING

Data Cleaning



@TheRecklessYes @japan_review @Mike1inFive ॐ, I must one day master Excel but that's why I had kids (I tell myself daily!) but I'm afraid I'm old school and just love paper. Though as a sop to the non-paper musicians I scan and convert now or use my Remarkable2 for converting written docs and drawings ॐµ€□ॐ«ॐ



Time to Leave... #PeterCapaldi #reMarkable2 #DoctorWho
<https://t.co/g8swL5KRPS>

Data Cleaning

```
def clean_tweets(sent):
    sent = sent.lower()
    sent = sent.encode('ascii', 'ignore').decode()      # encode to ascii unicode, it removes strange characters
    sent = re.sub(r'https\S+', '', sent)      # remove https
    sent = re.sub(r'http\S+', '', sent)      # remove http
    sent = re.sub(r'@\S+', '', sent)      # remove @
    sent = re.sub(r'#\S+', '', sent)      # remove #
    sent = " ".join(sent.split())
    return sent
```

FEATURE ENGINEERING

Feature Engineering



`Tweet.public_metrics: {'retweet_count': 76, 'reply_count': 1428, 'like_count': 4704, 'quote_count': 208}`



`User.public_metrics: {'followers_count': 727, 'following_count': 323, 'tweet_count': 86092, 'listed_count': 6}`

Feature Engineering

```
# get retweet, reply and like count from tweet.public_metrics
df['tweet.retweet_count'] = [re.findall(r'retwinet_count': (\d*), x)[0] for x in df['tweet.public_metrics']]

df['tweet.reply_count'] = [re.findall(r'reply_count': (\d*), x)[0] for x in df['tweet.public_metrics']]

df['tweet.like_count'] = [re.findall(r'like_count': (\d*), x)[0] for x in df['tweet.public_metrics']]

df['tweet.quote_count'] = [re.findall(r'quote_count': (\d*), x)[0] for x in df['tweet.public_metrics']]

# get retweet, reply and like count from user.public_metrics
df['user.followers_count'] = [re.findall(r'followers_count': (\d*), x)[0] for x in df['user.public_metrics']]

df['user.following_count'] = [re.findall(r'following_count': (\d*), x)[0] for x in df['user.public_metrics']]

df['user.tweet_count'] = [re.findall(r'tweet_count': (\d*), x)[0] for x in df['user.public_metrics']]

df['user.listed_count'] = [re.findall(r'listed_count': (\d*), x)[0] for x in df['user.public_metrics']]
```

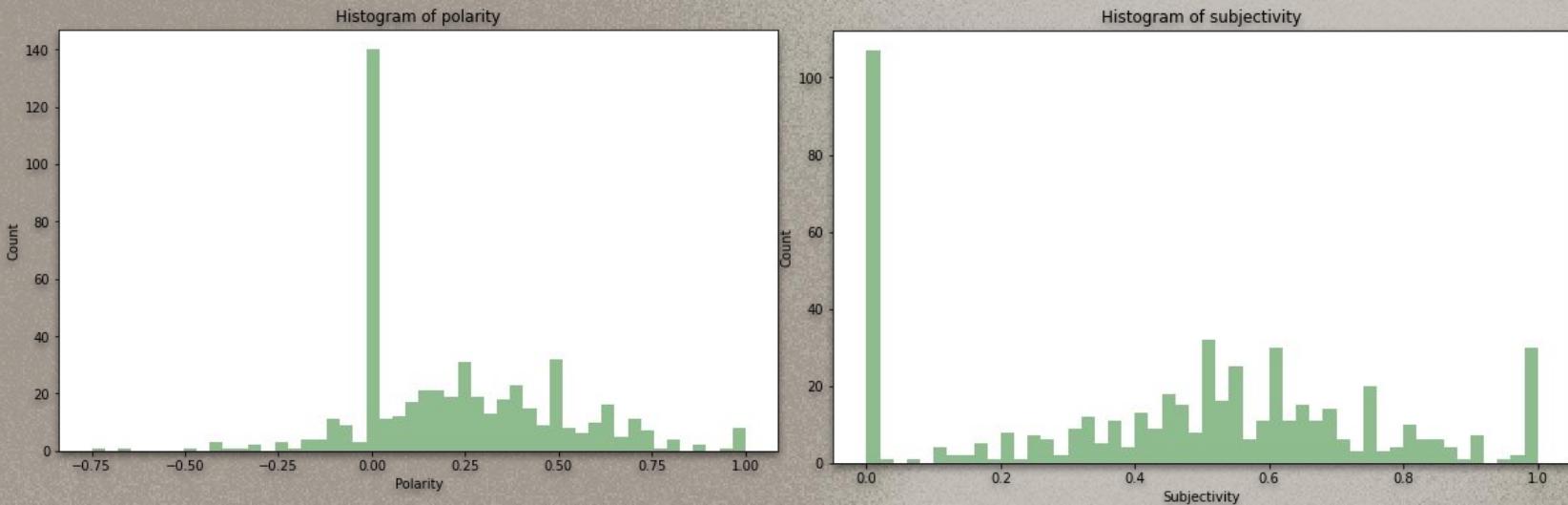
Feature Engineering

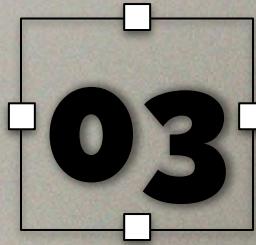
```
# sentiment analysis
def detect_sentiment(text):
    blob = TextBlob(text)
    return blob.sentiment.polarity

def detect_subjectivity(text):
    blob = TextBlob(text)
    return blob.sentiment.subjectivity

# create new columns for polarity and subjectivity of the reviews
df['polarity'] = df['tweet.text'].apply(detect_sentiment)
df['subjectivity'] = df['tweet.text'].apply(detect_subjectivity)
```

Feature Engineering

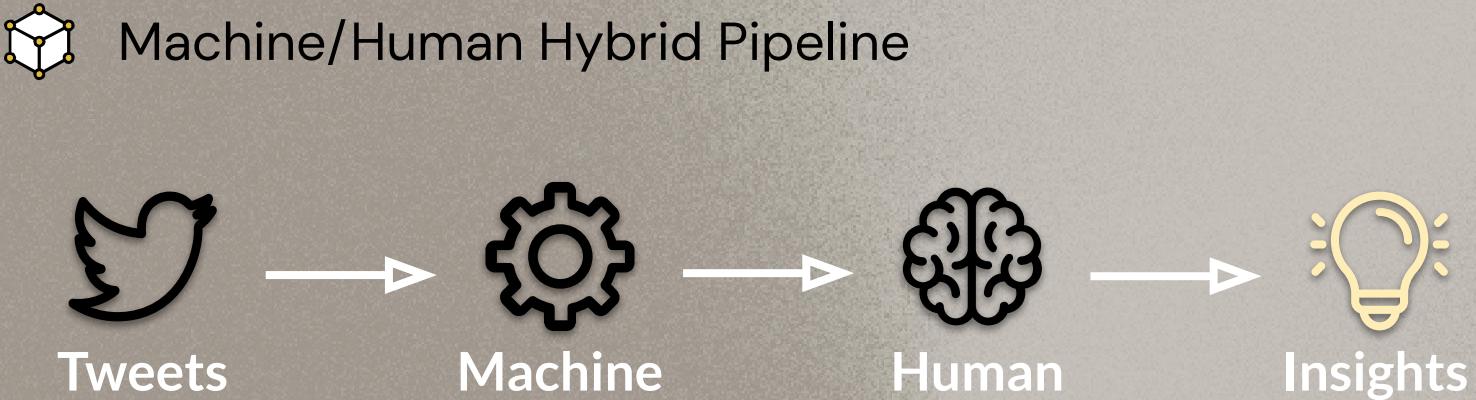




MODELING

Deriving insight from data

High level overview



High level overview



Inspired by methods backed by research



Paper: Identifying Customer Needs from UGC

Artem Timoshenko, John R. Hauser (2019) Marketing Science

38(1):1-20. <https://doi.org/10.1287/mksc.2018.1123>



Less expensive and more effective than
traditional marketing research

MACHINE

Machine: BERT Embedding



Filter Tweets



Sentence embedding

```

1 # this function uses simple BERT embedding
2 def simple_bert_embedding(data, model = 'all-MiniLM-L6-v2'):
3     # use model
4     embedder = SentenceTransformer(model)
5
6     # load sentences into corpus
7     corpus = data['tweet.text'].to_list()
8     corpus_embeddings = embedder.encode(corpus)
9
10    return corpus, corpus_embeddings

```

```

3 # get tweets that have more than 0 likes
4 tweets = tweets[tweets['tweet.like_count'] > 10]
5
6 # select only negative reviews to identify the problems
7 tweets = tweets[tweets['polarity'] < 0]
8 tweets.shape

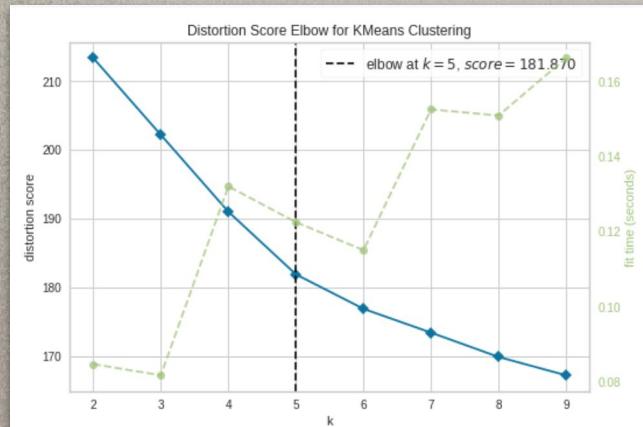
```

embeddings	tweet.text
[0.009987292, -0.110503115, -0.040178385, 0.02...]	my remarkable2 has not just changed the way i ...
[-0.046124715, -0.01279768, 0.023298763, -0.12...]	no matter what you think of - this video is re...



Clustering and number of topics

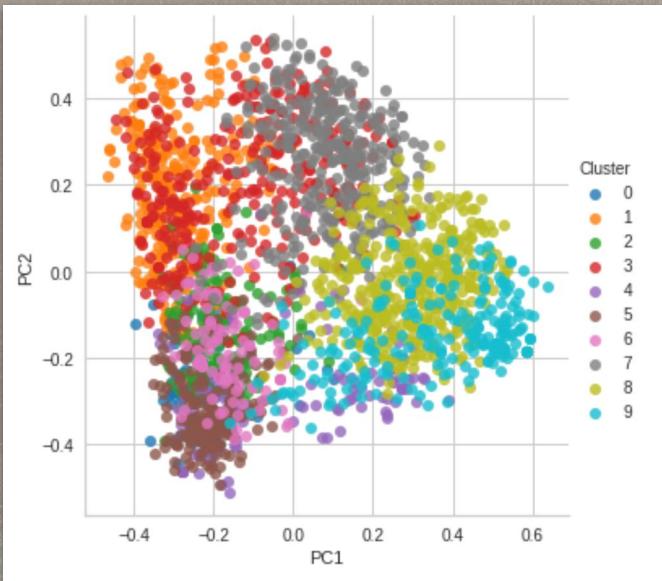
```
1 ## Import libraries
2 from nltk.cluster import KMeansClusterer
3 import nltk
4
5 import numpy as np
6 from scipy.spatial import distance_matrix
7
8 def clustering_tweets(data,NUM_CLUSTERS = 6):
9
10    sentences = data['tweet.text']
11
12    X = np.array(data['embeddings'].tolist())
13
14    kclusterer = KMeansClusterer(
15        NUM_CLUSTERS, distance=nltk.cluster.util.cosine_distance,
16        repeats=25, avoid_empty_clusters=True)
17
18    assigned_clusters = kclusterer.cluster(X, assign_clusters=True)
19
20    data['cluster'] = pd.Series(assigned_clusters, index=data.index)
21    data['centroid'] = data['cluster'].apply(lambda x: kclusterer.means()[x])
22
23    return data, assigned_clusters
24
25 clustering_tweets(tweets, 25)
```



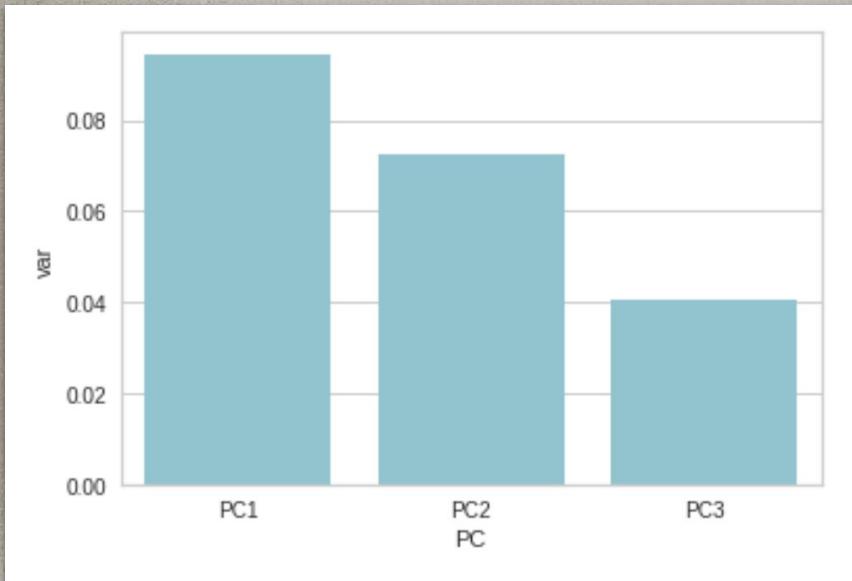
Visualizing Clusters



PCA: 348 → 3



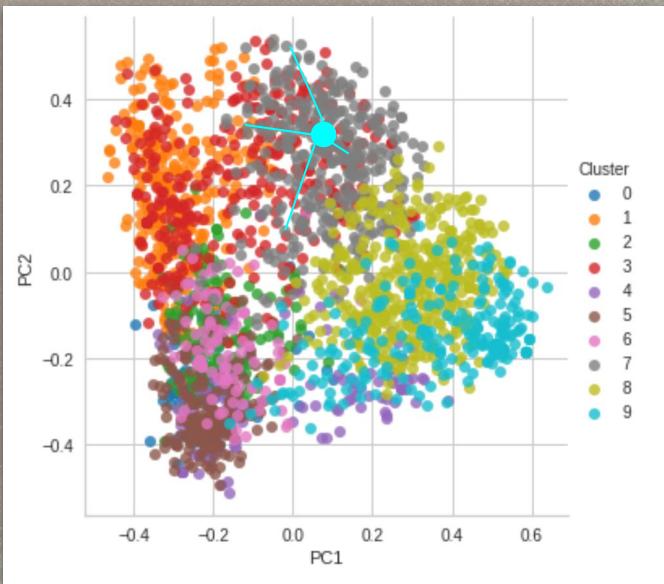
```
11 # load decomposition to do PCA analysis with sklearn
12 from sklearn import decomposition
13 pca = decomposition.PCA(n_components=3)
14 pc = pca.fit_transform(np.array(tweets['embeddings'].tolist()))
```



Top Tweets!



Distance from centroid



```

1 # calculating the distance from the centroid to find the most representative sentences
2 def distance_from_centroid(row):
3     # type of emb and centroid is different, hence using tolist below
4     return distance_matrix([row['embeddings']], [row['centroid']].tolist())[0][0]
5
6
7 # Compute centroid distance to the data
8 tweets['distance_from_centroid'] = tweets.apply(distance_from_centroid, axis=1)

```

	tweet.text	cluster	distance_from_centroid
44	the remarkable2 is worth it alone just to remi...	0	0.556198
14	spending the weekend cleaning my uncles apartm...	0	0.709915
58	i wish there was an economist & ft app for...	0	0.749116
92	was helping kiddo with precal homework. i got ...	0	0.752610
28	here it is my christmas present from main man....	1	0.695393
...
108	just receive my new geek device i badly want t...	9	0.816953
23	sometimes it's just easier to think without co...	9	0.869101
49	hello there!	9	0.935910
90	setting up a bbs over packet radio today to co...	9	0.956422
89	xiao's outfit was a struggle doodled on the	9	0.987239

Ready for interpretation



3 tweets per topic

	B	C	D
tweet.text		cluster	distance_from_centroid
2 i received a disappointing email from regarding the delayed shipping of my rema	0	0	0.554391007
3 issues with my remarkable2. contacted customer support. seem to be stuck in a	0	0	0.563296902
2 so my remarkable2 is expected to ship in a couple of days. been waiting since m	0	0	0.587904601
4 okay twitter. what do you think of the remarkable paper device? has terrible rev	1	1	0.735291335
5 buying the was probably one of the biggest changes of 2020 in my work routine.	1	1	0.741377045
9 impatiently waiting for the to start, with my expensive writing thingy i bought sp	1	1	0.753710607
0 6 weeks awaiting a refund, ignored support requests, bad product even worse s	2	2	0.687936129
9 unacceptable to have no live customer service. got tablet for wife for xmas, com	2	2	0.713093047
4 i was so disappointed in the lack of responsiveness of the i returned it in the sec	2	2	0.724430843
1 my tablet is randomly re opening notebooks, or other items when going into fol	3	3	0.718678107
7 for the last month, my rm1 stopped identifying wireless networks (of all sorts, a	3	3	0.748219296
0 my is sleeping and i cannot wake it up by any means. pushing power button for	3	3	0.758511449
1 i use a remarkable2 expensive but amazing if you want to write things down but	4	4	0.695966441
0 got my delivered yesterday, omg, i'm addicted to writing on it! it really does feel	4	4	0.698301397
6 have it (remarkable2) - and i love it. don't care much about transcribing. never r	4	4	0.740565766

HUMAN



Human interpretation

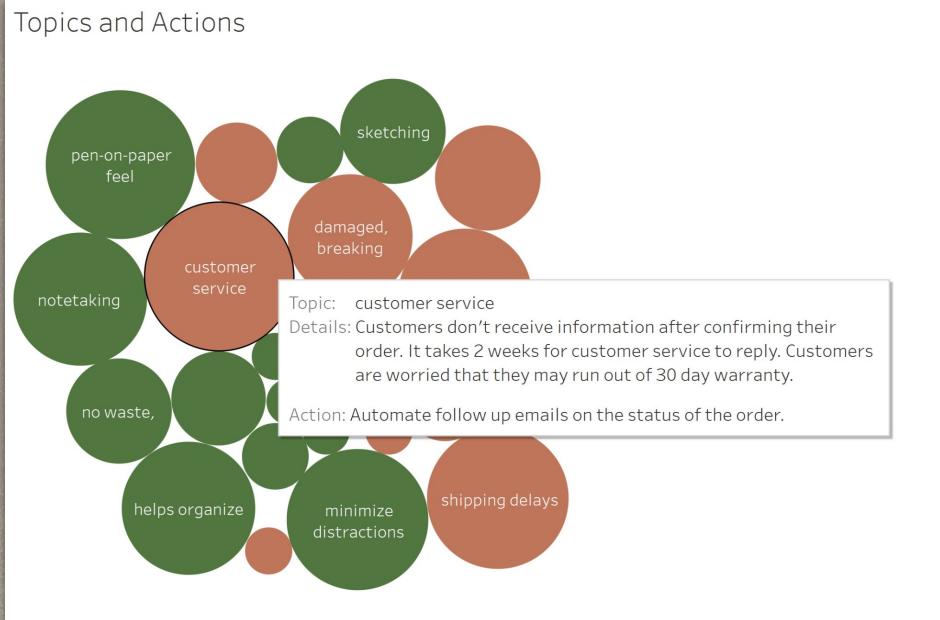
Topic, importance, polarity, details, action

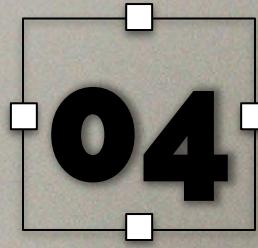
A	B	C	D	E	
1	Topic	Importance	Polarity	Details	Action
2	customer service	10	0	Customers don't receive information after confirming their order. It takes 2 weeks for customer service to reply. Customers are worried that they may run out of 30 day warranty.	Automate follow up emails on the status of the order.
3	shipping delays	9	0	Many negative comments and frustration on shipping delays.	Predict the number of sales for at least 6 months in advance. Stock enough inventory in the US so that you don't have to ship it from overseas. Anything longer than 1 week is unacceptable these days when people got used to Amazon Prime same day delivery.
4	high price	8	0	There were many tweets complaining about high prices.	Find ways to reduce the price of the product. Don't charge for the pen since that's the main function. Create a budget version of reMarkable2. reMarkable Core
5	damaged, breaking	7	0	Quality control issue. Based on the tweets some tablets arrive damaged or break in a few days.	Improve quality control. Arrange quick repair shops.
	can't refund, exchange	6	0	Customers received damaged products and couldn't refund or exchange. This adds a lot to people being reluctant to buy.	Make the refund and exchange very easy. You can collaborate with store chains such as Amazon, Ralphs, CVS where people drop it off

Intuitive visualization



Part of the final dashboard





04

ANALYSIS OUTCOME

Insights and recommendations

INSIGHTS



Positive Findings

-  Paper-like writing experience
-  Great for organizing notes
-  Minimize distractions
-  Custom Templates





1. Customer Service



No live customer service



Unresponsive, long wait time



Slow tech support

"guys, this is simply unacceptable. it's been over a week and i am still waiting for a reply..."





2. Wait time



Long wait for customer service



Long wait for shipping



Long wait for refund

"6 weeks awaiting a refund, ignored support requests ..."

"I received a disappointing email from regarding the delayed shipping of my reMarkable2. I was given a rather vague eta of 'the next few weeks' ..."



3. Pricing

- 💡 High price makes potential customer hesitate
- 💡 Satisfied customer also mention the high price
- 💡 Customer question the subscription plan

“... however still can't justify to myself the device's new lower price point, yet.”;

“a bit pricey, but well worth it”;

“you pay a premium for the hardware, but then find it is worthless without paying an additional \$8/mo ...”



4. Inconvenient File Export



Cloud not available without subscription



Cannot convert to text without subscription



Having to email files in order to save as text

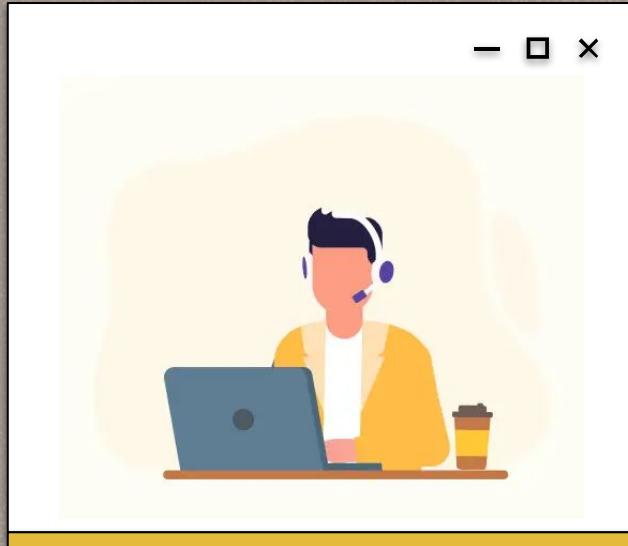
“... now, the cloud is unavailable, so I can't convert to text. btw, I can't save permanently to text without sending the file to someone via email anyway ...”



RECOMMENDATIONS



1. Better Customer Service



79%

66%



Live customer support



Live chat instead of phone support

- + Support multiple customers at once

- + Lower cost

- + Text records



2. Pricing & Hard/Software

	Price	RAM	Storage	File Type	Bluetooth
 reMarkable2	\$474 \$378 + \$8/mo	1 GB	8 GB	Pdf, ePub	NO
 Supernote A5X	\$459	2 GB	32 GB	Pdf, ePub, Kindle	YES
 Note Air 2	\$499	4 GB	32 GB	All types	YES

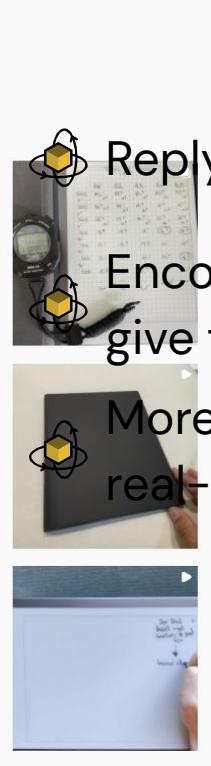
reMarkable @remarkablepaper Aug 3, 2021

We're thrilled to be supporting @TEDTalks for its #TEDMonterey, 'The Case for Optimism, conference. Check out the link to read how reMarkable helped bring #BetterThinking to this fantastic event and how we'll be part of creating #ideasworthspreading.

remarkable.com/blog/remarkabl...

12 4 24

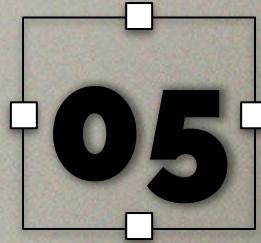
3. Twitter Presence



Reply to customer's tweets

Encourage customers to give feedback

More UGC, better real-time monitoring



DASHBOARD

Dashboards for reMarkable



Dashboard Functionalities



Visualize daily sentiment and insights from clusters



Manage sales data



View number of returned items



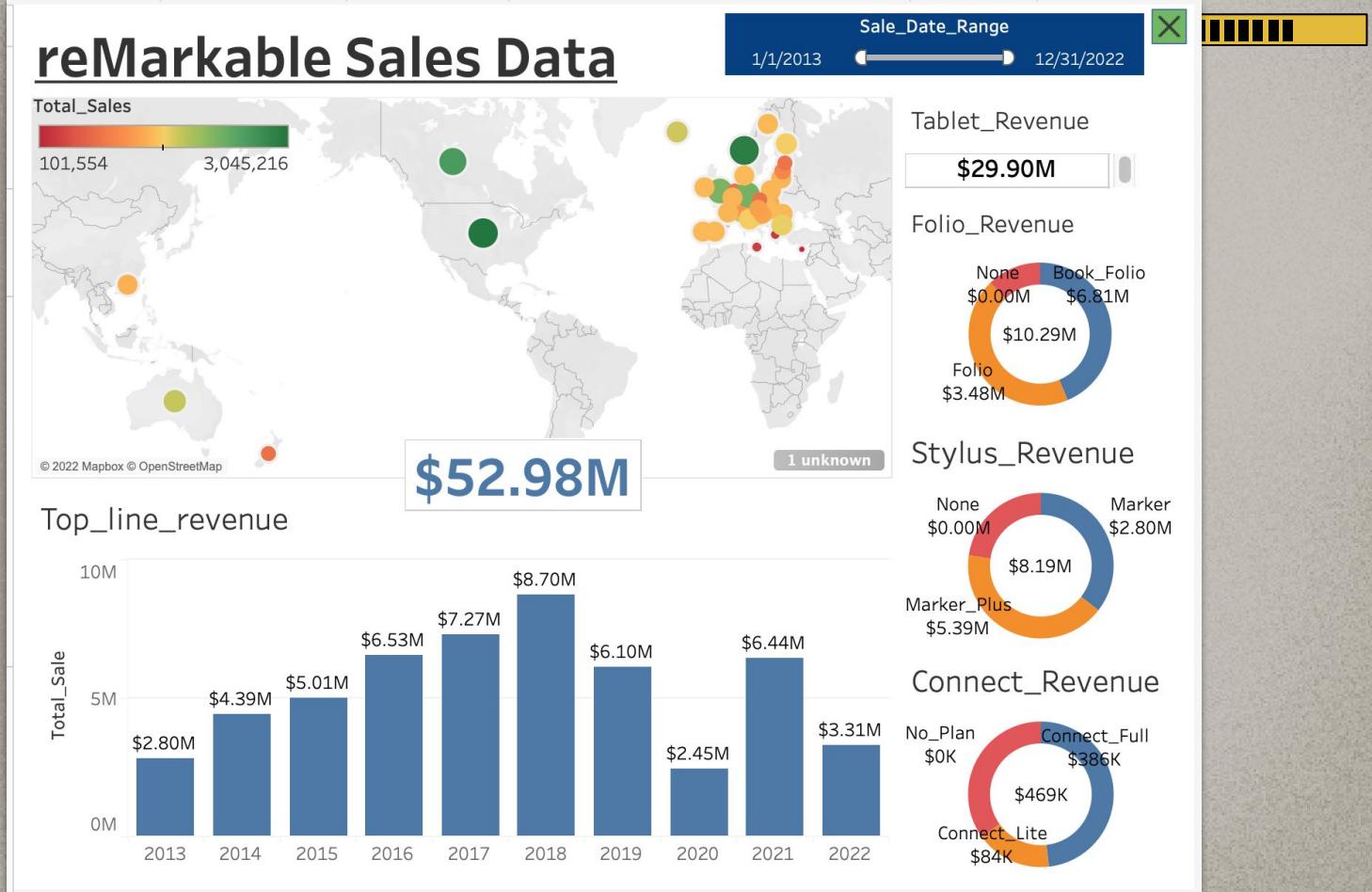
Assess customer ratings



Monitor customer service metrics,

* Data with regard to Revenue and Customer service are synthetic, and can be replace with real data with ease

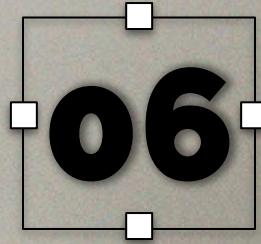




reMarkable Brand Monitoring Dashboard



[Link to Dashboard](#)



Further Action

Our next steps

Next Steps

-  Obtain useful review data from other sources (blog posts, etc.)
-  Obtain revenue data from reMarkable for further research
-  Discuss with Stakeholders



VP of Marketing



Sigurd Gran-Jansen



VP of Marketing and
Founding Employee



Project Meeting:

Monday (5/9) 10:00 AM



Open Questions



If you were to take notes on a tablet, what function would you like the tablet to have?



How much are you willing to pay for such a device?



Any questions you might have for the presentation



THANK YOU!

rM