Slack Data Analysis

The goal of this project is to develop a comprehensive data analysis and visualization pipeline for Slack data. The project involves a series of tasks that will enable us to gain insights, perform data science analysis, including topic modeling and sentiment analysis, create interactive dashboards for visualization and deployment. Let's dive into the details:

Dataset

Directory Layout

Directory Structure

```
data
└── anonymized
    ├── chang-w11
    ├── all-week3
    ├── all-week4
    ├── adludios-challange
    ├── all-week5
    ├── all-technical-support
    ├── all-week2
    ├── gokada-challenge-presentation
    ├── all-ideas
    ├── all-week11
    ├── all-week10
    ├── study-group
    ├── team-10
    ├── all-resources
    ├── week4-teamwork
    ├── all-ml-week12
    ├── all-career-exercises
    ├── data-engineering
    ├── week-2-group-8
```

Show Loading data

Show Loading data

	msg_type	msg_content	sender_nam
18	message	Yeah. Do you guys prefer Gmeet or Slack?	Robert Carte
19	message	or I can share my zoom, I find it doing well on a low internet connection	Anita Rodrig
20	message	For me, I am okay with either of those!	Anita Rodrig
21	message	zoom is fine with me too if it is possible	Phillip Atkin
22	message	Zoom is find by me. Please do share <@U03U9FWPNCE>	Robert Carte
23	message	what about <@U03UD4FEDHB> do you have zoom installed?	Phillip Atkin
24	message	Yaa!	Michael Gor
25	message	great!	Anita Rodrig
26	message	<https://cmu.zoom.us/j/93668953926?pwd=VDFpSDhQelBrdU9Dak95NFZ6S0hLQT09	Anita Rodrig
27	message	you can join when ready!	Anita Rodrig

Explanatory Data Analysis

Select Explanatory Data Analysis Option

Word Cloud

Generate Analysis

Prepare data for LDA Analysis

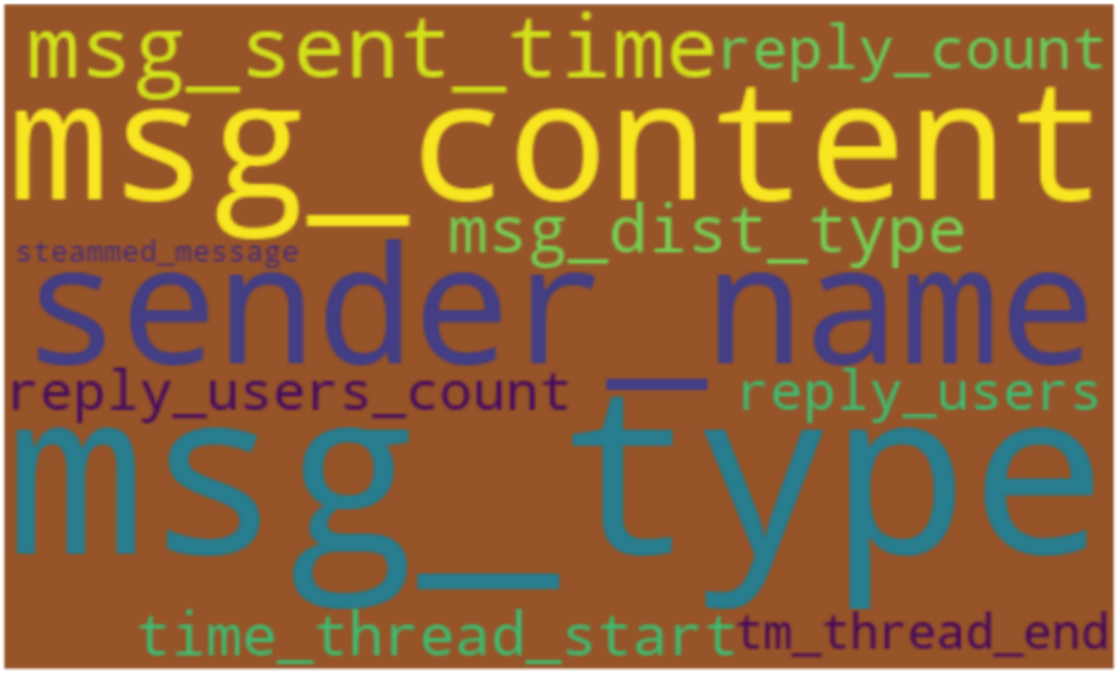
Explanatory Data Analysis

Select Explanatory Data Analysis Option

Word Cloud

Generate Analysis

WordCloud for week



Slack Data Analysis

The goal of this project is to develop a comprehensive data analysis and visualization pipeline for Slack data. The project involves a series of tasks that will enable us to gain insights, perform data science analysis, including topic modeling and sentiment analysis, create interactive dashboards for visualization and deployment. Let's dive into the details:

Dataset

Directory Layout

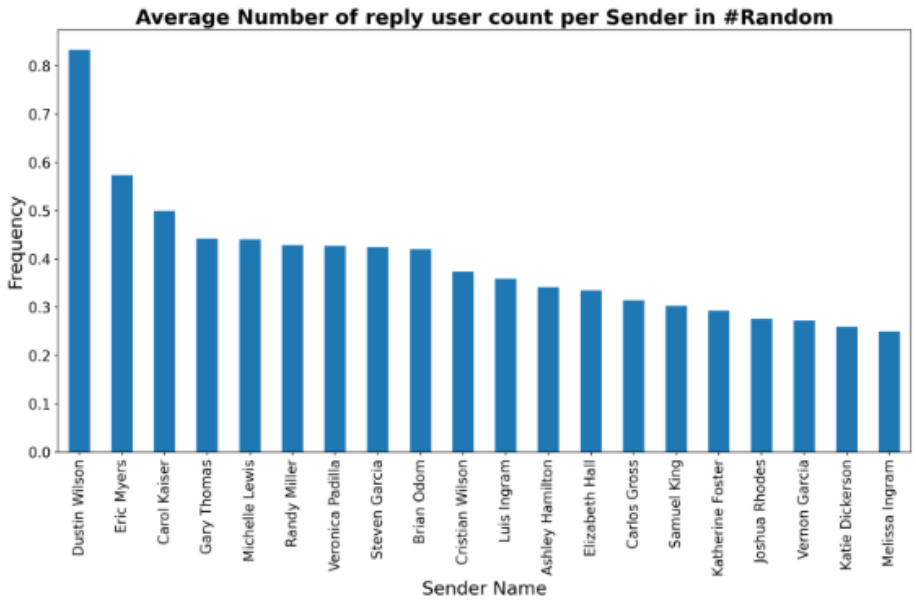
Show Loading data

Explanatory Data Analysis

Select Explanatory Data Analysis Option

Average Reply User Count

Generate Analysis



time_thread_starttm_thread_end

Prepare data for LDA Analysis

Vectorize the data

X_test_vectorized

Selected Vector: (0, 7506) 0.3780096471463705 (0, 6689) 0.4717127033850039 (0, 5357) 0.6077035436225605 (0, 3158) 0.5150652727548912 (1, 8029) 0.17363264195041445 (1, 7946) 0.2728348412654405 (1, 4632) 0.2894011020504597 (1, 4629) 0.1843200241231955 (1, 4433) 0.21953072602739065 (1, 4264) 0.3333107368292552 (1, 3391) 0.2275599984795108 (1, 3341) 0.22971575813229458 (1, 2614) 0.24430079423355194 (1, 1943) 0.28178710667612233 (1, 1238) 0.23904093995480108 (1, 826) 0.5635742133522447 (2, 8032) 0.29588198998271 (2, 7758) 0.3352307662173344 (2, 4708) 0.4545787507114176 (2, 2863) 0.5238625196059167 (2, 1351) 0.5648012051500932 (3, 7604) 0.35110526157481386 (3, 7373) 0.6904301423880937 (3, 5020) 0.6324802872631196 (4, 8062) 0.31959097978419043 : : (3546, 5123) 0.3886876897697207 (3546, 1106) 0.49632850432458253 (3547, 5919) 0.3416907423031333 (3547, 5556) 0.4806178027690066 (3547, 5393) 0.27581975058955893 (3547, 2838) 0.41680518242507264 (3547, 2658) 0.37828866338018813 (3547, 2572) 0.376123180472879 (3547, 1337) 0.34333643797758806 (3548, 7702) 0.37831350260667035 (3548, 7604) 0.25554826626266 (3548, 6729) 0.4377343728864681 (3548, 4380) 0.32335240575604896 (3548, 3592) 0.18253895185898908 (3548, 1810) 0.4298718044090603 (3548, 1379) 0.2013889413353188 (3548, 1266) 0.48655737069460525 (3549, 8444) 0.30127301656084493 (3549, 6846) 0.3619971893748483 (3549, 3843) 0.4229266961556911 (3549, 3246) 0.2794017376071221 (3549, 2962) 0.29658109905570795 (3549, 1011) 0.34434224257986323 (3549, 447) 0.4229266961556911 (3549, 69) 0.3685937661095237

Topic modeling

Trained LDA Model:

LdaModel<num_terms=8854, num_topics=5, decay=0.5, chunksize=2000>

(0, '0.025*"ye" + 0.015*"tri" + 0.008*"think" + 0.008*"man" + 0.008*"smile" + 0.007*"read" + 0.006*"face" + 0.006*"one" + 0.006*"data" + 0.005*"lol"')

(1, '0.027*"yeah" + 0.017*"ok" + 0.015*"hi" + 0.011*"group" + 0.011*"lie" + 0.009*"join" + 0.009*"link" + 0.008*"pleas" + 0.008*"anyon" + 0.007*"ujkjgraq"')

(2, '0.027*"joy" + 0.019*"laugh" + 0.014*"floor" + 0.014*"roll" + 0.013*"grin" + 0.011*"sure" +

Prepare data for LDA Analysis

Vectorize the data

X_train_vectorized

Selected Vector: `(0, 1149) 0.14968358498178186 (0, 7054) 0.2021423056276625 (0, 6828) 0.398677484740141 (0, 7152) 0.1375924370372861 (0, 863) 0.24601955792251334 (0, 8032) 0.11736773946378581 (0, 2832) 0.2126339526876437 (0, 4971) 0.22253133407678505 (0, 6117) 0.2981440427370504 (0, 4459) 0.17608543825742293 (0, 355) 0.37663415652754045 (0, 1592) 0.2633093909750301 (0, 2236) 0.40141427123754037 (0, 1383) 0.16815875805796077 (0, 3859) 0.2633093909750301 (1, 3571) 0.43874048791109044 (1, 4246) 0.43829566804160497 (1, 8733) 0.43829566804160497 (1, 1379) 0.27161428678035643 (1, 3144) 0.38365833521434717 (1, 4739) 0.376460420233424 (1, 3592) 0.24619121035181668 (2, 384) 0.38044017541571157 (2, 813) 0.40597604544499755 (2, 955) 0.32847100039571503 : : (14194, 7375) 0.5174405178512552 (14194, 4045) 0.450769091498227 64 (14195, 7247) 0.7237536709838115 (14195, 2726) 0.38270830948333256 (14195, 6343) 0.3434275507946725 (14195, 5447) 0.313041551723047 (14195, 4904) 0.337309172481073 (14196, 1315) 0.5352859244648458 (14196, 274) 0.43719672738518683 (14196, 7892) 0.47949752080488883 (14196, 6738) 0.4242963105489059 (14196, 4442) 0.33523539346263787 (14197, 2764) 0.33690390859870367 (14197, 69) 0.330009267342598 (14197, 6651) 0.3555581982414162 (14197, 2572) 0.2515530671993697 (14197, 5880) 0.25450231517605504 (14197, 2658) 0.2530013530152467 (14197, 8309) 0.268287386913315 (14197, 5940) 0.24924677391337577 (14197, 8029) 0.21001856174618289 (14197, 5393) 0.18446963084736467 (14197, 4904) 0.2340355757758858 (14197, 8445) 0.33968491116192223 (14197, 293) 0.2769666802371285`

Topic modeling

Trained LDA Model:

`LdaModel<num_terms=8854, num_topics=5, decay=0.5, chunksize=2000>`

`(0, '0.022*"thank" + 0.014*"meet" + 0.014*"yeah" + 0.013*"laugh" + 0.010*"floor" + 0.010*"roll" + 0.009*"googl" + 0.008*"time" + 0.008*"sure" + 0.007*"join"')`

`(1, '0.025*"good" + 0.021*"morn" + 0.012*"smile" + 0.011*"grin" + 0.008*"face" + 0.007*"challeng"`