```
import pandas as pd

df1=pd.read_csv('input.csv')

df=pd.read_csv('input.csv')
data = df[:1]['Content']

len(df)
```

```
401
```

```
from langchain_groq import ChatGroq
import os
os.environ['GROQ_API_KEY'] =
'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'

llm = ChatGroq(
    model="llama3-70b-8192",
    temperature=0,
    max_tokens=None,
    timeout=None,
    max_retries=2,
)

from langchain_core.prompts import PromptTemplate

prompt=PromptTemplate.from_template(
    """ Scrapted text from website:
    {data}
    Instruction: Please Provide information of the incident in brief
way:"""
)

prompt2=PromptTemplate.from_template(
    """ Write the gist of the {data} in one line."""
)

llm_chain=prompt|llm

response=llm_chain.invoke(input={"data":data})
print(response)
```

```
content='Here is a brief summary of the incident:\n\nSix people were
killed and 30 others injured in an incident.' additional_kwargs={}
response_metadata={'token_usage': {'completion_tokens': 23,
'prompt_tokens': 53, 'total_tokens': 76, 'completion_time':
0.065714286, 'prompt_time': 0.001418165, 'queue_time': 0.055437221,
'total_time': 0.067132451}, 'model_name': 'llama3-70b-8192',
'system_fingerprint': 'fp_dd4ae1c591', 'finish_reason': 'stop',
'logprobs': None} id='run-e295e26a-d7ee-434f-b05c-b6fae6507a89-0'
```

```
usage_metadata={'input_tokens': 53, 'output_tokens': 23,
'total_tokens': 76}
```

```
# !pip install gensim

# !pip install nltk

import pandas as pd
from gensim import corpora, models
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk

nltk.download('punkt')
nltk.download('stopwords')


df = pd.read_csv("input.csv")


df['Combined_Text'] = df[['Headline', 'Content']].fillna('').agg('
'.join, axis=1)


texts = df['Combined_Text'].apply(
    lambda x: [word.lower() for word in word_tokenize(x) if
word.isalpha() and word.lower() not in stopwords.words('english')]
)


dictionary = corpora.Dictionary(texts)
corpus = [dictionary.doc2bow(text) for text in texts]


lda_model = models.LdaModel(corpus, num_topics=5, id2word=dictionary,
passes=10)


for idx, topic in lda_model.print_topics(-1):
    print(f"Topic {idx + 1}: {topic}")
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\palla\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\palla\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
Topic 1: 0.016*"police" + 0.009*"said" + 0.008*"bus" + 0.006*"road" +
0.006*"news" + 0.006*"fire" + 0.006*"accident" + 0.006*"students" +
0.005*"also" + 0.005*"area"
Topic 2: 0.019*"road" + 0.017*"said" + 0.011*"bus" + 0.011*"news" +
```

```
0.011*"police" + 0.009*"killed" + 0.009*"accidents" + 0.009*"daily" +
0.009*"star" + 0.007*"dhaka"
Topic 3: 0.010*"said" + 0.008*"road" + 0.007*"news" + 0.006*"accident"
+ 0.006*"blood" + 0.005*"people" + 0.005*"daily" + 0.005*"star" +
0.004*"transport" + 0.004*"police"
Topic 4: 0.022*"injured" + 0.020*"said" + 0.020*"police" +
0.020*"killed" + 0.018*"upazila" + 0.017*"bus" + 0.015*"news" +
0.013*"road" + 0.013*"people" + 0.012*"around"
Topic 5: 0.020*"police" + 0.015*"said" + 0.014*"hospital" +
0.014*"bus" + 0.012*"news" + 0.009*"injured" + 0.009*"dhaka" +
0.008*"around" + 0.008*"road" + 0.007*"star"
```

```
# !pip install pyLDAvis

import pyLDAvis.gensim_models as gensimvis
import pyLDAvis

pyLDAvis.enable_notebook()
vis = gensimvis.prepare(lda_model, corpus, dictionary)
pyLDAvis.display(vis)
```

```
<IPython.core.display.HTML object>
```

```
for i, row in enumerate(lda_model[corpus]):
    print(f"Document {i}: Topic Distribution {row}")
```

```
Document 0: Topic Distribution [(3, 0.99214125)]
Document 1: Topic Distribution [(4, 0.9951398)]
Document 2: Topic Distribution [(3, 0.99308664)]
Document 3: Topic Distribution [(4, 0.9903712)]
Document 4: Topic Distribution [(3, 0.9363636), (4, 0.058426734)]
Document 5: Topic Distribution [(2, 0.44262585), (3, 0.5536131)]
Document 6: Topic Distribution [(3, 0.99076915)]
Document 7: Topic Distribution [(1, 0.6933626), (3, 0.29969415)]
Document 8: Topic Distribution [(1, 0.046849288), (3, 0.9509271)]
Document 9: Topic Distribution [(3, 0.80679905), (4, 0.19206586)]
Document 10: Topic Distribution [(1, 0.0282333), (3, 0.96987057)]
Document 11: Topic Distribution [(3, 0.9919262)]
Document 12: Topic Distribution [(4, 0.9947338)]
Document 13: Topic Distribution [(1, 0.9928964)]
Document 14: Topic Distribution [(1, 0.99276686)]
Document 15: Topic Distribution [(3, 0.9976875)]
Document 16: Topic Distribution [(1, 0.5619039), (3, 0.4367241)]
Document 17: Topic Distribution [(0, 0.9982276)]
Document 18: Topic Distribution [(1, 0.9938056)]
Document 19: Topic Distribution [(4, 0.99586475)]
Document 20: Topic Distribution [(1, 0.9184214), (3, 0.08014692)]
Document 21: Topic Distribution [(3, 0.6295198), (4, 0.36658224)]
Document 22: Topic Distribution [(0, 0.10030721), (4, 0.8944047)]
Document 23: Topic Distribution [(0, 0.2583509), (3, 0.7347594)]
```

```
Document 24: Topic Distribution [(0, 0.11799243), (3, 0.79703826), (4,
0.07889296)]
Document 25: Topic Distribution [(3, 0.992809)]
Document 26: Topic Distribution [(0, 0.9986358)]
Document 27: Topic Distribution [(0, 0.28066123), (3, 0.7104624)]
Document 28: Topic Distribution [(1, 0.43683794), (3, 0.55431604)]
Document 29: Topic Distribution [(1, 0.8325495), (3, 0.16662207)]
Document 30: Topic Distribution [(1, 0.767202), (3, 0.22497746)]
Document 31: Topic Distribution [(1, 0.11479267), (3, 0.8813969)]
Document 32: Topic Distribution [(3, 0.9931301)]
Document 33: Topic Distribution [(1, 0.9928599)]
Document 34: Topic Distribution [(1, 0.9977675)]
Document 35: Topic Distribution [(3, 0.21447536), (4, 0.7830709)]
Document 36: Topic Distribution [(3, 0.4023511), (4, 0.59380794)]
Document 37: Topic Distribution [(1, 0.5474858), (3, 0.45120183)]
Document 38: Topic Distribution [(1, 0.9934432)]
Document 39: Topic Distribution [(2, 0.93598676), (3, 0.059038993)]
Document 40: Topic Distribution [(1, 0.9410748), (3, 0.052355766)]
Document 41: Topic Distribution [(1, 0.14397089), (3, 0.8511136)]
Document 42: Topic Distribution [(3, 0.99509335)]
Document 43: Topic Distribution [(3, 0.97886837), (4, 0.017890386)]
Document 44: Topic Distribution [(3, 0.13264671), (4, 0.8620566)]
Document 45: Topic Distribution [(3, 0.9709879), (4, 0.023177328)]
Document 46: Topic Distribution [(1, 0.10098215), (3, 0.89667636)]
Document 47: Topic Distribution [(1, 0.5895309), (3, 0.40495774)]
Document 48: Topic Distribution [(2, 0.23577715), (3, 0.7595696)]
Document 49: Topic Distribution [(3, 0.52906716), (4, 0.46886936)]
Document 50: Topic Distribution [(1, 0.9930465)]
Document 51: Topic Distribution [(1, 0.9900658)]
Document 52: Topic Distribution [(1, 0.011718467), (3, 0.9847324)]
Document 53: Topic Distribution [(3, 0.4788643), (4, 0.51636356)]
Document 54: Topic Distribution [(1, 0.67554414), (3, 0.32323253)]
Document 55: Topic Distribution [(1, 0.28913888), (3, 0.15676346), (4,
0.5534082)]
Document 56: Topic Distribution [(0, 0.42273694), (3, 0.57599366)]
Document 57: Topic Distribution [(1, 0.27531642), (2, 0.7233035)]
Document 58: Topic Distribution [(1, 0.6517352), (3, 0.23325633), (4,
0.11395076)]
Document 59: Topic Distribution [(3, 0.9975879)]
Document 60: Topic Distribution [(1, 0.5201516), (3, 0.47667122)]
Document 61: Topic Distribution [(2, 0.32780194), (3, 0.66505206)]
Document 62: Topic Distribution [(0, 0.9972706)]
Document 63: Topic Distribution [(1, 0.9931942)]
Document 64: Topic Distribution [(1, 0.9987883)]
Document 65: Topic Distribution [(3, 0.54094434), (4, 0.4510416)]
Document 66: Topic Distribution [(1, 0.5645321), (3, 0.42515984)]
Document 67: Topic Distribution [(1, 0.9457052), (2, 0.04641601)]
Document 68: Topic Distribution [(1, 0.5685079), (3, 0.17328079), (4,
0.2530841)]
```

```
Document 69: Topic Distribution [(1, 0.99451447)]
Document 70: Topic Distribution [(1, 0.998415)]
Document 71: Topic Distribution [(1, 0.9949003)]
Document 72: Topic Distribution [(1, 0.20524323), (3, 0.79314727)]
Document 73: Topic Distribution [(1, 0.6704494), (3, 0.3283394)]
Document 74: Topic Distribution [(3, 0.99580246)]
Document 75: Topic Distribution [(3, 0.9916966)]
Document 76: Topic Distribution [(3, 0.9930789)]
Document 77: Topic Distribution [(1, 0.9954302)]
Document 78: Topic Distribution [(3, 0.99424386)]
Document 79: Topic Distribution [(1, 0.34398177), (2, 0.043987233),
(3, 0.6062678)]
Document 80: Topic Distribution [(1, 0.48518306), (2, 0.3578524), (4,
0.15410008)]
Document 81: Topic Distribution [(1, 0.30718496), (3, 0.68346703)]
Document 82: Topic Distribution [(1, 0.067759074), (3, 0.9227575)]
Document 83: Topic Distribution [(1, 0.9926687)]
Document 84: Topic Distribution [(1, 0.99458605)]
Document 85: Topic Distribution [(1, 0.99713427)]
Document 86: Topic Distribution [(0, 0.994673)]
Document 87: Topic Distribution [(1, 0.24033037), (3, 0.75443274)]
Document 88: Topic Distribution [(1, 0.86788845), (3, 0.122538514)]
Document 89: Topic Distribution [(4, 0.99571335)]
Document 90: Topic Distribution [(3, 0.99544454)]
Document 91: Topic Distribution [(3, 0.75542235), (4, 0.23931734)]
Document 92: Topic Distribution [(3, 0.99341935)]
Document 93: Topic Distribution [(3, 0.9928595)]
Document 94: Topic Distribution [(3, 0.9893776)]
Document 95: Topic Distribution [(3, 0.08116164), (4, 0.9153057)]
Document 96: Topic Distribution [(0, 0.21894619), (3, 0.7756387)]
Document 97: Topic Distribution [(3, 0.99681556)]
Document 98: Topic Distribution [(3, 0.98907)]
Document 99: Topic Distribution [(1, 0.26920357), (3, 0.72471166)]
Document 100: Topic Distribution [(0, 0.99356174)]
Document 101: Topic Distribution [(3, 0.98496675)]
Document 102: Topic Distribution [(4, 0.98785466)]
Document 103: Topic Distribution [(3, 0.98975354)]
Document 104: Topic Distribution [(0, 0.99349433)]
Document 105: Topic Distribution [(3, 0.9904832)]
Document 106: Topic Distribution [(3, 0.769041), (4, 0.22464727)]
Document 107: Topic Distribution [(3, 0.2958112), (4, 0.6943878)]
Document 108: Topic Distribution [(3, 0.9934852)]
Document 109: Topic Distribution [(3, 0.99195856)]
Document 110: Topic Distribution [(1, 0.99758244)]
Document 111: Topic Distribution [(3, 0.9890264)]
Document 112: Topic Distribution [(3, 0.4949074), (4, 0.4985366)]
Document 113: Topic Distribution [(1, 0.9154768), (3, 0.08194311)]
Document 114: Topic Distribution [(1, 0.99636555)]
Document 115: Topic Distribution [(3, 0.45264068), (4, 0.5449327)]
```

```
Document 116: Topic Distribution [(3, 0.99680823)]
Document 117: Topic Distribution [(3, 0.22996454), (4, 0.7630344)]
Document 118: Topic Distribution [(3, 0.9893497)]
Document 119: Topic Distribution [(3, 0.99434555)]
Document 120: Topic Distribution [(3, 0.99165606)]
Document 121: Topic Distribution [(0, 0.88642895), (3, 0.11146453)]
Document 122: Topic Distribution [(3, 0.9949662)]
Document 123: Topic Distribution [(1, 0.45164585), (2, 0.54421085)]
Document 124: Topic Distribution [(1, 0.994948)]
Document 125: Topic Distribution [(3, 0.9980124)]
Document 126: Topic Distribution [(3, 0.9974009)]
Document 127: Topic Distribution [(4, 0.9946885)]
Document 128: Topic Distribution [(1, 0.20309061), (3, 0.7958072)]
Document 129: Topic Distribution [(3, 0.99811363)]
Document 130: Topic Distribution [(3, 0.9887923)]
Document 131: Topic Distribution [(3, 0.99157596)]
Document 132: Topic Distribution [(1, 0.09332754), (2, 0.45308512),
(3, 0.4525087)]
Document 133: Topic Distribution [(2, 0.10628791), (3, 0.8918419)]
Document 134: Topic Distribution [(1, 0.9911694)]
Document 135: Topic Distribution [(3, 0.9951949)]
Document 136: Topic Distribution [(1, 0.43293807), (3, 0.55807954)]
Document 137: Topic Distribution [(1, 0.9942016)]
Document 138: Topic Distribution [(2, 0.80395114), (3, 0.19205275)]
Document 139: Topic Distribution [(1, 0.09446519), (3, 0.020086892),
(4, 0.8822745)]
Document 140: Topic Distribution [(3, 0.98908585)]
Document 141: Topic Distribution [(1, 0.8954325), (4, 0.098124735)]
Document 142: Topic Distribution [(3, 0.9955393)]
Document 143: Topic Distribution [(3, 0.99054414)]
Document 144: Topic Distribution [(1, 0.98899716)]
Document 145: Topic Distribution [(1, 0.39124385), (3, 0.602686)]
Document 146: Topic Distribution [(4, 0.9923476)]
Document 147: Topic Distribution [(2, 0.89305484), (4, 0.10089226)]
Document 148: Topic Distribution [(2, 0.9976162)]
Document 149: Topic Distribution [(3, 0.9935253)]
Document 150: Topic Distribution [(3, 0.45759714), (4, 0.53819925)]
Document 151: Topic Distribution [(3, 0.64791185), (4, 0.34556428)]
Document 152: Topic Distribution [(1, 0.3017408), (3, 0.69271255)]
Document 153: Topic Distribution [(3, 0.7668753), (4, 0.22818345)]
Document 154: Topic Distribution [(1, 0.63595647), (3, 0.15390167),
(4, 0.20552552)]
Document 155: Topic Distribution [(3, 0.9911542)]
Document 156: Topic Distribution [(4, 0.9940136)]
Document 157: Topic Distribution [(3, 0.7627881), (4, 0.22866848)]
Document 158: Topic Distribution [(2, 0.7501263), (3, 0.24738823)]
Document 159: Topic Distribution [(1, 0.99115026)]
Document 160: Topic Distribution [(1, 0.35816032), (3, 0.6380415)]
Document 161: Topic Distribution [(1, 0.34514916), (3, 0.6510106)]
```

```
Document 162: Topic Distribution [(3, 0.6781027), (4, 0.31589502)]
Document 163: Topic Distribution [(1, 0.06576788), (2, 0.84196377),
(3, 0.09105487)]
Document 164: Topic Distribution [(3, 0.986029)]
Document 165: Topic Distribution [(1, 0.42179447), (3, 0.57037795)]
Document 166: Topic Distribution [(1, 0.79786485), (3, 0.19557784)]
Document 167: Topic Distribution [(1, 0.18766916), (3, 0.8071434)]
Document 168: Topic Distribution [(1, 0.99433094)]
Document 169: Topic Distribution [(1, 0.96032286), (3, 0.038031016)]
Document 170: Topic Distribution [(3, 0.9935267)]
Document 171: Topic Distribution [(3, 0.7922618), (4, 0.2004935)]
Document 172: Topic Distribution [(4, 0.99409837)]
Document 173: Topic Distribution [(2, 0.49039274), (3, 0.50603426)]
Document 174: Topic Distribution [(2, 0.3602642), (3, 0.6323553)]
Document 175: Topic Distribution [(3, 0.26415133), (4, 0.7284456)]
Document 176: Topic Distribution [(1, 0.75899607), (3, 0.23485483)]
Document 177: Topic Distribution [(3, 0.99139434)]
Document 178: Topic Distribution [(3, 0.15233661), (4, 0.8436679)]
Document 179: Topic Distribution [(3, 0.9875027)]
Document 180: Topic Distribution [(3, 0.99649227)]
Document 181: Topic Distribution [(1, 0.9964701)]
Document 182: Topic Distribution [(0, 0.22201891), (4, 0.7701368)]
Document 183: Topic Distribution [(1, 0.9716059), (3, 0.024407938)]
Document 184: Topic Distribution [(0, 0.56871986), (3, 0.42516777)]
Document 185: Topic Distribution [(1, 0.9959747)]
Document 186: Topic Distribution [(3, 0.9904593)]
Document 187: Topic Distribution [(1, 0.713635), (3, 0.28044683)]
Document 188: Topic Distribution [(0, 0.13790253), (1, 0.38761055),
(3, 0.46897486)]
Document 189: Topic Distribution [(4, 0.9884984)]
Document 190: Topic Distribution [(1, 0.57122046), (3, 0.42551607)]
Document 191: Topic Distribution [(4, 0.9973126)]
Document 192: Topic Distribution [(4, 0.9954829)]
Document 193: Topic Distribution [(4, 0.98850065)]
Document 194: Topic Distribution [(4, 0.99802464)]
Document 195: Topic Distribution [(4, 0.998015)]
Document 196: Topic Distribution [(3, 0.989487)]
Document 197: Topic Distribution [(3, 0.5430286), (4, 0.4505209)]
Document 198: Topic Distribution [(1, 0.46416217), (3, 0.5284279)]
Document 199: Topic Distribution [(3, 0.99005604)]
Document 200: Topic Distribution [(3, 0.98675567)]
Document 201: Topic Distribution [(3, 0.8873529), (4, 0.10695556)]
Document 202: Topic Distribution [(1, 0.23865351), (3, 0.7553135)]
Document 203: Topic Distribution [(3, 0.47569072), (4, 0.5156094)]
Document 204: Topic Distribution [(1, 0.99590963)]
Document 205: Topic Distribution [(4, 0.99824595)]
Document 206: Topic Distribution [(0, 0.3065382), (3, 0.68418145)]
Document 207: Topic Distribution [(2, 0.14675778), (3, 0.8502937)]
Document 208: Topic Distribution [(3, 0.78151155), (4, 0.20913824)]
```

```
Document 209: Topic Distribution [(1, 0.883333), (3, 0.11016898)]
Document 210: Topic Distribution [(1, 0.92764986), (3, 0.067496255)]
Document 211: Topic Distribution [(0, 0.30388576), (3, 0.68653643)]
Document 212: Topic Distribution [(3, 0.6582737), (4, 0.3324676)]
Document 213: Topic Distribution [(4, 0.99025977)]
Document 214: Topic Distribution [(2, 0.18976432), (3, 0.8027259)]
Document 215: Topic Distribution [(1, 0.44161767), (3, 0.5537946)]
Document 216: Topic Distribution [(1, 0.6672964), (3, 0.32985523)]
Document 217: Topic Distribution [(3, 0.99148756)]
Document 218: Topic Distribution [(1, 0.99742305)]
Document 219: Topic Distribution [(1, 0.44297504), (3, 0.5489033)]
Document 220: Topic Distribution [(1, 0.9926374)]
Document 221: Topic Distribution [(1, 0.32493058), (3, 0.6661124)]
Document 222: Topic Distribution [(1, 0.99121267)]
Document 223: Topic Distribution [(1, 0.124546275), (3, 0.867745)]
Document 224: Topic Distribution [(2, 0.33239222), (3, 0.6594328)]
Document 225: Topic Distribution [(3, 0.9914639)]
Document 226: Topic Distribution [(1, 0.06522555), (3, 0.9290273)]
Document 227: Topic Distribution [(3, 0.99265116)]
Document 228: Topic Distribution [(3, 0.9886353)]
Document 229: Topic Distribution [(0, 0.16192117), (1, 0.5487334), (2,
0.056149133), (3, 0.23138405)]
Document 230: Topic Distribution [(3, 0.985702)]
Document 231: Topic Distribution [(1, 0.9948447)]
Document 232: Topic Distribution [(3, 0.9864875)]
Document 233: Topic Distribution [(1, 0.9957512)]
Document 234: Topic Distribution [(3, 0.9902524)]
Document 235: Topic Distribution [(3, 0.9482462), (4, 0.04552095)]
Document 236: Topic Distribution [(3, 0.9929098)]
Document 237: Topic Distribution [(3, 0.9923361)]
Document 238: Topic Distribution [(3, 0.99685854)]
Document 239: Topic Distribution [(1, 0.19513535), (3, 0.7984448)]
Document 240: Topic Distribution [(1, 0.50204253), (3, 0.31357175),
(4, 0.18129854)]
Document 241: Topic Distribution [(3, 0.9903454)]
Document 242: Topic Distribution [(1, 0.9889615)]
Document 243: Topic Distribution [(3, 0.23822634), (4, 0.753317)]
Document 244: Topic Distribution [(3, 0.9926436)]
Document 245: Topic Distribution [(3, 0.34649163), (4, 0.64442897)]
Document 246: Topic Distribution [(0, 0.3704445), (3, 0.6209665)]
Document 247: Topic Distribution [(3, 0.15579778), (4, 0.83388317)]
Document 248: Topic Distribution [(1, 0.99334806)]
Document 249: Topic Distribution [(3, 0.46678984), (4, 0.5245935)]
Document 250: Topic Distribution [(1, 0.986673)]
Document 251: Topic Distribution [(3, 0.99021953)]
Document 252: Topic Distribution [(1, 0.9978277)]
Document 253: Topic Distribution [(1, 0.52811366), (3, 0.4656985)]
Document 254: Topic Distribution [(1, 0.7289721), (3, 0.2659212)]
Document 255: Topic Distribution [(1, 0.94077724), (3, 0.05035669)]
```

```
Document 256: Topic Distribution [(1, 0.3651781), (3, 0.63201094)]
Document 257: Topic Distribution [(3, 0.4873211), (4, 0.50890034)]
Document 258: Topic Distribution [(3, 0.9897377)]
Document 259: Topic Distribution [(3, 0.7181782), (4, 0.2773301)]
Document 260: Topic Distribution [(1, 0.46632978), (3, 0.14050743),
(4, 0.38822722)]
Document 261: Topic Distribution [(3, 0.9915744)]
Document 262: Topic Distribution [(3, 0.8139749), (4, 0.177331)]
Document 263: Topic Distribution [(3, 0.33017367), (4, 0.6631547)]
Document 264: Topic Distribution [(1, 0.99388385)]
Document 265: Topic Distribution [(3, 0.987765)]
Document 266: Topic Distribution [(0, 0.36182988), (3, 0.52520865),
(4, 0.10974978)]
Document 267: Topic Distribution [(1, 0.99431425)]
Document 268: Topic Distribution [(0, 0.4435688), (3, 0.54900676)]
Document 269: Topic Distribution [(1, 0.29892367), (3, 0.6990144)]
Document 270: Topic Distribution [(1, 0.8022646), (3, 0.19284938)]
Document 271: Topic Distribution [(3, 0.98582244)]
Document 272: Topic Distribution [(3, 0.3351693), (4, 0.6592478)]
Document 273: Topic Distribution [(3, 0.98835045)]
Document 274: Topic Distribution [(2, 0.045695655), (3, 0.94692826)]
Document 275: Topic Distribution [(3, 0.9919376)]
Document 276: Topic Distribution [(1, 0.26314467), (3, 0.43348187),
(4, 0.2986887)]
Document 277: Topic Distribution [(3, 0.9925043)]
Document 278: Topic Distribution [(2, 0.9937273)]
Document 279: Topic Distribution [(3, 0.98651075)]
Document 280: Topic Distribution [(1, 0.98995155)]
Document 281: Topic Distribution [(3, 0.5326807), (4, 0.46311724)]
Document 282: Topic Distribution [(3, 0.99249107)]
Document 283: Topic Distribution [(1, 0.22318289), (3, 0.65118074),
(4, 0.12219624)]
Document 284: Topic Distribution [(1, 0.99122405)]
Document 285: Topic Distribution [(1, 0.30524778), (3, 0.626935), (4,
0.06358644)]
Document 286: Topic Distribution [(0, 0.5854342), (1, 0.40995458)]
Document 287: Topic Distribution [(1, 0.6455813), (3, 0.08683984), (4,
0.26082507)]
Document 288: Topic Distribution [(3, 0.39847612), (4, 0.59806496)]
Document 289: Topic Distribution [(3, 0.9914207)]
Document 290: Topic Distribution [(3, 0.5287024), (4, 0.4649206)]
Document 291: Topic Distribution [(1, 0.61863524), (3, 0.37167582)]
Document 292: Topic Distribution [(2, 0.9973906)]
Document 293: Topic Distribution [(1, 0.7870052), (3, 0.2079202)]
Document 294: Topic Distribution [(3, 0.9930602)]
Document 295: Topic Distribution [(3, 0.98774683)]
Document 296: Topic Distribution [(3, 0.9894848)]
Document 297: Topic Distribution [(4, 0.9935168)]
Document 298: Topic Distribution [(1, 0.29775012), (3, 0.69587785)]
```

```
Document 299: Topic Distribution [(3, 0.98808986)]
Document 300: Topic Distribution [(1, 0.8470913), (3, 0.14535497)]
```

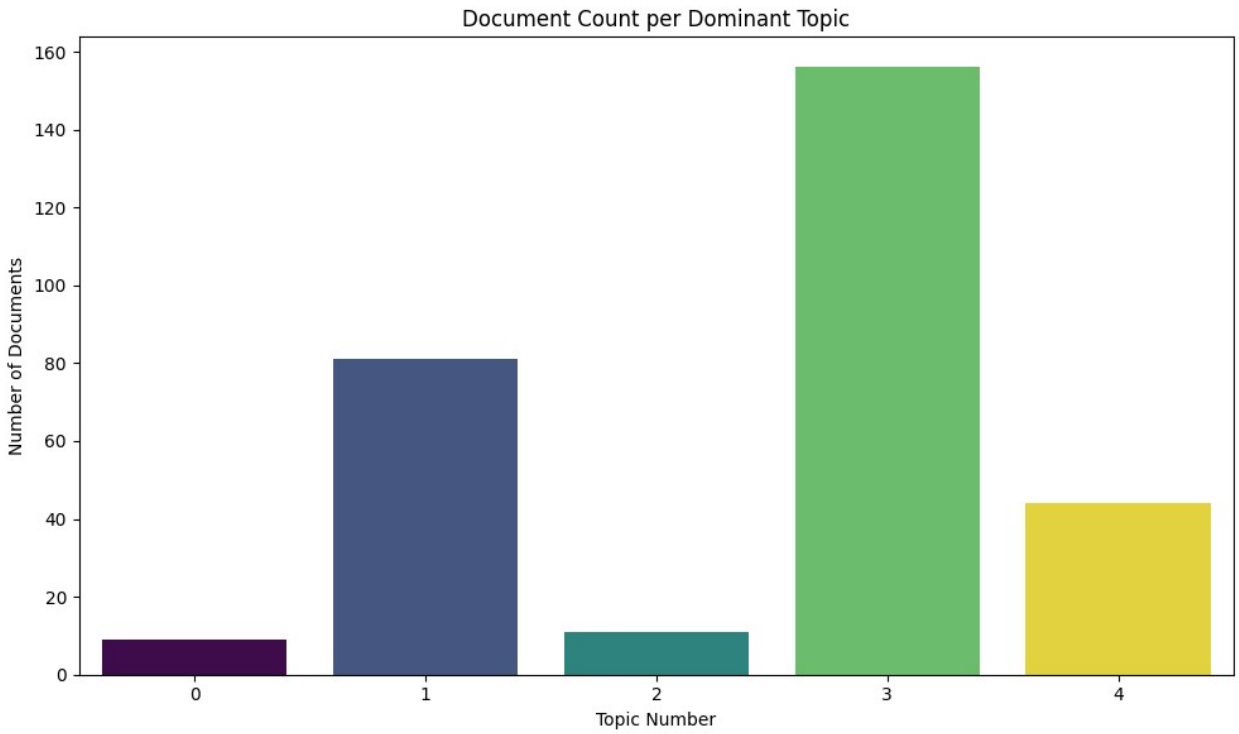
chart that shows which document belongs to which dominant topic

```
import pandas as pd


doc_topics = []
for i, row in enumerate(lda_model[corpus]):
    row = sorted(row, key=lambda x: (x[1]), reverse=True)
    topic_num, prop_topic = row[0]
    doc_topics.append((i, int(topic_num), round(prop_topic, 4)))

df_topic_doc = pd.DataFrame(doc_topics, columns=['Document_No',
'Dominant_Topic', 'Topic_Perc_Contrib'])

df_topic_doc = pd.concat([df_topic_doc, df[['Headline', 'Content']]],
axis=1)

# !pip install seaborn

print(df_topic_doc[['Document_No', 'Dominant_Topic']])
```

```
     Document_No  Dominant_Topic
0              0               3
1              1               4
2              2               3
3              3               4
4              4               3
..           ...             ...
296          296               3
297          297               4
298          298               3
299          299               3
300          300               1

[301 rows x 2 columns]
```

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10,6))
sns.countplot(data=df_topic_doc, x='Dominant_Topic',
hue='Dominant_Topic', palette='viridis', legend=False)
plt.title('Document Count per Dominant Topic')
plt.xlabel('Topic Number')
plt.ylabel('Number of Documents')
plt.tight_layout()
```

```
plt.savefig('dominant_topic_distribution.png', dpi=300)

plt.show()
```




```
prompt_template = """
You are an expert in topic modeling. Given the top keywords of a topic
from LDA, assign a short and meaningful name that best represents the
topic.

Keywords: {keywords}

Topic Name:
"""

from langchain.prompts import PromptTemplate
from langchain.chains import LLMChain


topics = [
    ['police', 'said', 'bus', 'road', 'news', 'fire', 'accident',
'students', 'also', 'area'],
    ['road', 'said', 'bus', 'news', 'police', 'killed', 'accidents',
```

```
'daily', 'star', 'dhaka'],
    ['said', 'road', 'news', 'accident', 'blood', 'people', 'daily',
'star', 'transport', 'police'],
    ['injured', 'said', 'police', 'killed', 'upazila', 'bus', 'news',
'road', 'people', 'around'],
    ['police', 'said', 'hospital', 'bus', 'news', 'injured', 'dhaka',
'around', 'road', 'star']
]



prompt_template = PromptTemplate(
    input_variables=["topic_keywords"],
    template="Given the following keywords from a topic model:\
n{topic_keywords}\n\nGive a short and descriptive innovative name for
this topic so that no two topic name coincide for news paper."
)

llm_chain = LLMChain(llm=llm, prompt=prompt_template)

for i, topic_keywords in enumerate(topics, start=1):
    result = llm_chain.run(topic_keywords=topic_keywords)
    print(f"Topic {i} name: {result}")
```

```
Topic 1 name: Based on the keywords, I would suggest the following
innovative topic name:

**"Safety Watch"**

This name captures the essence of the topic, which appears to be
related to incidents, accidents, and safety concerns in a specific
area, involving authorities like the police and fire department. The
name is short, descriptive, and unique, making it unlikely to coincide
with other topic names in a newspaper.
Topic 2 name: Based on the keywords, I would suggest the following
topic name:

**"Wheels of Tragedy"**

This name captures the essence of the topic, which appears to be
related to road accidents, bus crashes, and police involvement,
resulting in fatalities. The phrase "Wheels of Tragedy" is short,
memorable, and evocative, conveying a sense of danger and loss. It's
also unique and unlikely to coincide with other topic names in a
newspaper.
Topic 3 name: Based on the keywords, I would suggest the following
topic name:

**"StreetBeat"**
```

This name captures the essence of the topic, which appears to be related to news, accidents, and transportation. The word "Street" conveys a sense of location and setting, while "Beat" implies a rhythm or pulse, suggesting a dynamic and active topic. The name is also catchy and easy to remember, making it a great fit for a newspaper topic.

No two topic names are likely to coincide with "StreetBeat", as it's a unique and creative combination of words that effectively summarizes the theme of the topic.
Topic 4 name: Based on the keywords, I would suggest the following topic name:

"Road to Tragedy"

This name captures the essence of the topic, which appears to be related to accidents or incidents involving buses on roads, resulting in injuries or fatalities. The name is short, descriptive, and evocative, making it suitable for a news article or topic.
Topic 5 name: Based on the keywords, I would suggest the following topic name:

"Urban Crisis"

This name captures the essence of the topic, which appears to be related to accidents, injuries, and incidents involving transportation (bus, road) and authorities (police) in an urban setting (Dhaka). The name is short, descriptive, and innovative, making it unlikely to coincide with other topic names in a newspaper.

```
import pandas as pd
from gensim import corpora, models
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk
import matplotlib.pyplot as plt

nltk.download('punkt')
nltk.download('stopwords')

df = pd.read_csv("input.csv")

df['Combined_Text'] = df[['Headline', 'Content']].fillna('').agg('
'.join, axis=1)

df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

texts = df['Combined_Text'].apply(
    lambda x: [word.lower() for word in word_tokenize(x) if
```

```
word.isalpha() and word.lower() not in stopwords.words('english')]
)

dictionary = corpora.Dictionary(texts)
corpus = [dictionary.doc2bow(text) for text in texts]

lda_model = models.LdaModel(corpus, num_topics=5, id2word=dictionary,
passes=15)


topic_distributions = []
for doc in corpus:
    topic_probs = lda_model.get_document_topics(doc)


    print(topic_probs)


    topic_probs_dict = {}
    for prob in topic_probs:

        topic_probs_dict[f"Topic_{prob[0] + 1}"] = prob[1]

    topic_distributions.append(topic_probs_dict)

topic_df = pd.DataFrame(topic_distributions)


topic_df['Date'] = df['Date']


topic_trends =
topic_df.groupby(topic_df['Date'].dt.to_period('M')).mean()


plt.figure(figsize=(10, 6))
for topic in topic_trends.columns:
    if topic != 'Date':
        plt.plot(topic_trends.index.astype(str), topic_trends[topic],
label=topic)

plt.title('Topic Trends Over Time')
plt.xlabel('Date')
plt.ylabel('Topic Proportion')
plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()

plt.savefig('TimeSeriesAccident.png', dpi=300)

plt.show()
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\palla\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\palla\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!

[(1, 0.99211526)]
[(1, 0.4032356), (4, 0.5931325)]
[(1, 0.6277458), (4, 0.36709115)]
[(1, 0.7988564), (4, 0.19399707)]
[(1, 0.9931043)]
[(1, 0.99492687)]
[(1, 0.6596432), (2, 0.333467)]
[(1, 0.990806)]
[(0, 0.19351533), (1, 0.80426764)]
[(1, 0.998491)]
[(1, 0.99748915)]
[(1, 0.99198836)]
[(1, 0.99476373)]
[(0, 0.99282795)]
[(3, 0.71479523), (4, 0.27971798)]
[(1, 0.99769396)]
[(1, 0.6192971), (3, 0.37932882)]
[(0, 0.9982329)]
[(0, 0.9938074)]
[(0, 0.99076855)]
[(0, 0.012793973), (1, 0.85711056), (4, 0.12913525)]
[(1, 0.99481684)]
[(1, 0.41592088), (4, 0.5788032)]
[(1, 0.9909145)]
[(1, 0.987942)]
[(1, 0.99281794)]
[(4, 0.99863476)]
[(1, 0.98830056)]
[(1, 0.68149614), (4, 0.30968472)]
[(1, 0.98431814), (4, 0.014846024)]
[(1, 0.80665225), (3, 0.18556625)]
[(1, 0.80713195), (3, 0.18905903)]
[(1, 0.99314314)]
[(1, 0.99278367)]
[(1, 0.033029385), (4, 0.96529156)]
[(1, 0.9967439)]
[(1, 0.99487776)]
[(1, 0.998253)]
[(0, 0.9934094)]
[(0, 0.8645742), (1, 0.13045037)]
[(1, 0.8508848), (3, 0.14254802)]
[(1, 0.99346817)]
[(1, 0.86953324), (4, 0.12677765)]
```

```
[(1, 0.9956803)]
[(1, 0.99295056)]
[(1, 0.99223137)]
[(1, 0.99690783)]
[(1, 0.64447576), (3, 0.3500416)]
[(1, 0.9938149)]
[(1, 0.9972527)]
[(0, 0.9930138)]
[(1, 0.2258204), (3, 0.766719)]
[(1, 0.99528146)]
[(1, 0.81416285), (4, 0.1810444)]
[(1, 0.6737411), (2, 0.3250307)]
[(1, 0.6915475), (3, 0.30741763)]
[(1, 0.73382336), (4, 0.26491424)]
[(0, 0.9981493)]
[(1, 0.54921925), (3, 0.4491858)]
[(1, 0.9975934)]
[(1, 0.9957739)]
[(1, 0.99042755)]
[(3, 0.73758096), (4, 0.26037386)]
[(0, 0.70624197), (3, 0.28861642)]
[(0, 0.887931), (1, 0.111159384)]
[(1, 0.9893578)]
[(0, 0.41829962), (1, 0.5713939)]
[(4, 0.9895084)]
[(1, 0.80332404), (3, 0.1889485)]
[(1, 0.031418864), (4, 0.9644796)]
[(3, 0.9984068)]
[(2, 0.21852767), (3, 0.77760965)]
[(1, 0.9978418)]
[(1, 0.86842144), (3, 0.130361)]
[(1, 0.99580187)]
[(1, 0.9916996)]
[(1, 0.9930981)]
[(0, 0.3108179), (4, 0.68574977)]
[(1, 0.63366884), (2, 0.36205116)]
[(1, 0.9884544)]
[(0, 0.37161756), (1, 0.45602274), (4, 0.16951817)]
[(1, 0.9874977)]
[(1, 0.98719805)]
[(0, 0.22726975), (4, 0.7672095)]
[(1, 0.32740602), (4, 0.6685505)]
[(4, 0.99712723)]
[(4, 0.9946978)]
[(1, 0.99309736)]
[(1, 0.27170163), (4, 0.71881366)]
[(0, 0.84630686), (1, 0.1504722)]
[(1, 0.99544376)]
[(1, 0.894694), (4, 0.100054234)]
```

```
[(1, 0.99342275)]
[(1, 0.99288166)]
[(1, 0.98939085)]
[(1, 0.9952873)]
[(1, 0.9928459)]
[(1, 0.23539427), (2, 0.76224357)]
[(1, 0.69947743), (2, 0.29233155)]
[(1, 0.8179004), (3, 0.17604518)]
[(0, 0.99357885)]
[(1, 0.94646853), (4, 0.042183504)]
[(1, 0.6620176), (3, 0.32888603)]
[(1, 0.9897892)]
[(4, 0.99353814)]
[(1, 0.9904729)]
[(1, 0.9916018)]
[(1, 0.98696154)]
[(1, 0.9934949)]
[(1, 0.9919287)]
[(2, 0.9975699)]
[(1, 0.8244586), (4, 0.16730525)]
[(1, 0.7605826), (4, 0.23285232)]
[(1, 0.10509441), (2, 0.8923295)]
[(3, 0.9963466)]
[(1, 0.74093837), (2, 0.25663862)]
[(1, 0.9968079)]
[(1, 0.52120835), (4, 0.47179276)]
[(1, 0.9893779)]
[(1, 0.9943534)]
[(1, 0.9916857)]
[(1, 0.1637621), (2, 0.83414453)]
[(1, 0.9949836)]
[(2, 0.60397667), (3, 0.3918522)]
[(3, 0.99494636)]
[(1, 0.99801743)]
[(1, 0.77914304), (2, 0.2189142)]
[(2, 0.99422115)]
[(1, 0.91241676), (3, 0.08647774)]
[(1, 0.99810964)]
[(1, 0.9888149)]
[(1, 0.9915763)]
[(1, 0.268404), (2, 0.7299762)]
[(2, 0.9973262)]
[(0, 0.99110806)]
[(1, 0.54963547), (2, 0.44678152)]
[(1, 0.8163231), (4, 0.17474914)]
[(0, 0.99416786)]
[(1, 0.3135795), (4, 0.68242896)]
[(1, 0.29883367), (4, 0.6964039)]
[(1, 0.98908734)]
```

```
[(1, 0.9914144)]
[(1, 0.9955403)]
[(1, 0.9904806)]
[(4, 0.9889646)]
[(1, 0.9919538)]
[(1, 0.18257084), (4, 0.8116834)]
[(1, 0.068450846), (2, 0.34057721), (3, 0.5869415)]
[(2, 0.9976042)]
[(1, 0.49440265), (2, 0.500748)]
[(1, 0.99436796)]
[(1, 0.99129426)]
[(1, 0.895003), (3, 0.09946216)]
[(1, 0.9934006)]
[(0, 0.3718596), (1, 0.62118036)]
[(1, 0.99121857)]
[(1, 0.3010031), (3, 0.69450897)]
[(1, 0.9885648)]
[(1, 0.88603336), (4, 0.111477114)]
[(0, 0.64255667), (1, 0.3508084)]
[(1, 0.8663491), (3, 0.12986049)]
[(4, 0.9948854)]
[(1, 0.9920012)]
[(0, 0.77610517), (1, 0.22207472)]
[(1, 0.9860798)]
[(1, 0.8602664), (3, 0.13196392)]
[(1, 0.46726775), (4, 0.526185)]
[(1, 0.99315494)]
[(3, 0.99429)]
[(1, 0.2359176), (4, 0.7624262)]
[(0, 0.49749315), (1, 0.49766743)]
[(1, 0.99035954)]
[(1, 0.53487265), (4, 0.46070117)]
[(1, 0.9952245)]
[(1, 0.607926), (4, 0.38471422)]
[(1, 0.9103769), (3, 0.08220953)]
[(1, 0.4333778), (4, 0.5604703)]
[(1, 0.9913921)]
[(1, 0.9946772)]
[(1, 0.9875354)]
[(1, 0.99650884)]
[(0, 0.91417325), (1, 0.08318119)]
[(0, 0.28001016), (1, 0.7130901)]
[(0, 0.62262565), (1, 0.37338752)]
[(1, 0.5874395), (4, 0.40647298)]
[(0, 0.36860228), (1, 0.026814332), (3, 0.4257264), (4, 0.17785358)]
[(1, 0.99047947)]
[(0, 0.15501441), (1, 0.8391268)]
[(1, 0.98895633)]
[(1, 0.98857903)]
```

```
[(1, 0.9956687)]
[(4, 0.9973063)]
[(1, 0.4584586), (4, 0.53816694)]
[(1, 0.5617209), (4, 0.42968118)]
[(3, 0.9980226)]
[(3, 0.9980139)]
[(1, 0.98953414)]
[(1, 0.9914338)]
[(1, 0.7540666), (3, 0.23851223)]
[(1, 0.9901433)]
[(1, 0.98678267)]
[(1, 0.99243796)]
[(3, 0.9919748)]
[(1, 0.81498224), (2, 0.17635588)]
[(2, 0.48973706), (3, 0.5071624)]
[(1, 0.3529313), (2, 0.64578956)]
[(1, 0.8951999), (4, 0.09560393)]
[(1, 0.9731318), (2, 0.02392618)]
[(1, 0.9875777)]
[(1, 0.9913917)]
[(1, 0.74031013), (4, 0.25484934)]
[(1, 0.76864636), (4, 0.22187972)]
[(1, 0.9877183)]
[(0, 0.5061717), (1, 0.48653576)]
[(1, 0.72379404), (2, 0.2687249)]
[(1, 0.99385995)]
[(1, 0.7611738), (2, 0.068892345), (4, 0.16804029)]
[(1, 0.99149936)]
[(0, 0.99742115)]
[(1, 0.98919886)]
[(1, 0.21483877), (4, 0.77967423)]
[(1, 0.9880715)]
[(0, 0.34929818), (1, 0.6441661)]
[(1, 0.98979425)]
[(0, 0.2543739), (1, 0.7375323)]
[(1, 0.8827782), (3, 0.1108178)]
[(1, 0.6040288), (2, 0.39017212)]
[(1, 0.9926785)]
[(1, 0.9886546)]
[(1, 0.26059592), (4, 0.7340074)]
[(1, 0.9857857)]
[(3, 0.99482816)]
[(1, 0.9865255)]
[(1, 0.17325537), (2, 0.8235715)]
[(1, 0.9902888)]
[(1, 0.8758044), (4, 0.118041836)]
[(1, 0.83185995), (3, 0.16280355)]
[(1, 0.9923509)]
[(1, 0.9968753)]
```

```
[(1, 0.9914861)]
[(1, 0.9937499)]
[(1, 0.9903556)]
[(1, 0.9889336)]
[(1, 0.80270886), (4, 0.18887654)]
[(1, 0.99266165)]
[(1, 0.98791486)]
[(1, 0.7517795), (4, 0.23967735)]
[(1, 0.9862973)]
[(4, 0.99334306)]
[(1, 0.9885767)]
[(0, 0.98663276)]
[(1, 0.99022317)]
[(3, 0.99782234)]
[(1, 0.99178755)]
[(0, 0.9932452)]
[(0, 0.4710534), (1, 0.253314), (3, 0.2697487)]
[(1, 0.9962293)]
[(1, 0.8381066), (4, 0.15809612)]
[(1, 0.9897806)]
[(1, 0.8631654), (3, 0.13233356)]
[(0, 0.2539435), (1, 0.7386369)]
[(1, 0.9916006)]
[(1, 0.9884282)]
[(1, 0.75513333), (4, 0.23821475)]
[(3, 0.043680895), (4, 0.9517349)]
[(1, 0.9877974)]
[(1, 0.7966758), (3, 0.1985385)]
[(2, 0.99426764)]
[(0, 0.5012592), (1, 0.49137965)]
[(1, 0.99725634)]
[(1, 0.71156967), (3, 0.28354743)]
[(1, 0.9469311), (3, 0.04241469)]
[(1, 0.7401137), (3, 0.25430235)]
[(1, 0.9884848)]
[(1, 0.9902204)]
[(1, 0.9919076)]
[(1, 0.9906158)]
[(1, 0.8311248), (4, 0.16325423)]
[(4, 0.9937479)]
[(1, 0.78853476), (3, 0.20137343)]
[(3, 0.9898847)]
[(1, 0.67531073), (4, 0.3204953)]
[(1, 0.99250484)]
[(1, 0.9930784)]
[(0, 0.11760862), (1, 0.74762815), (4, 0.13038488)]
[(1, 0.9914453)]
[(4, 0.99382186)]
[(0, 0.24999905), (1, 0.7398642)]
```

```
[(0, 0.2226912), (1, 0.7738416)]
[(1, 0.99142945)]
[(1, 0.9914929)]
[(1, 0.655843), (4, 0.3345004)]
[(4, 0.9973863)]
[(1, 0.44232056), (4, 0.55260646)]
[(1, 0.9930891)]
[(1, 0.98778343)]
[(1, 0.98953164)]
[(0, 0.7864749), (1, 0.2086729)]
[(1, 0.9915253)]
[(1, 0.9881366)]
[(0, 0.28942963), (1, 0.70304745)]
```