```
# Install required libraries
!pip install google-search-results
```

```
Requirement already satisfied: google-search-results in
/usr/local/lib/python3.10/dist-packages (2.4.2)
Requirement already satisfied: requests in
/usr/local/lib/python3.10/dist-packages (from google-search-results)
(2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->google-search-
results) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests->google-search-
results) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->google-search-
results) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->google-search-
results) (2024.8.30)
```

Data Collection: Collect Articles using SerpAPI

```
from serpapi import GoogleSearch
import pandas as pd

# Define SerpApi API key
api_key =
"[redacted]"

# Define queries for Air India's transformation by aspect
queries_and_topics = [
    # Operational Efficiency
    ("Operational Efficiency",
     {
        "pre": [
            "Air India operational inefficiencies before 2022",
            "Challenges in Air India's flight delays and route
planning before privatization",
            "How did Air India's operational costs compare to
competitors before 2022?"
        ],
        "post": [
            "Air India operational improvements post privatization
2022..2024",
            "Operational cost reductions at Air India under Tata Group
2022..2024"
        ]
     }),
```

```
    # Market Share
    ("Market Share",
     {
        "pre": [
            "Air India market share challenges in the domestic
aviation market before 2022",
            "Competition faced by Air India from IndiGo and SpiceJet
before privatization",
            "What was Air India’s international market share before
privatization in 2022?"
        ],
        "post": [
            "Air India market share growth post privatization
2022..2024",
            "Market positioning strategies for Air India under Tata
Group 2022..2024"
        ]
     }),
    # Financial Performance
    ("Financial Performance",
     {
        "pre": [
            "Air India financial struggles before 2022",
            "Debt burden and financial inefficiencies at Air India
before privatization",
            "Attempts to reduce Air India's operational losses before
2022"
        ],
        "post": [
            "Air India financial turnaround post privatization
2022..2024",
            "Profitability and cost-cutting measures at Air India
under Tata Group"
        ]
     }),
    # Customer Experience
    ("Customer Experience",
     {
        "pre": [
            "Air India customer satisfaction issues before 2022",
            "What were the major complaints about Air India's services
before privatization?",
            "Comparison of Air India's customer reviews with
competitors before 2022"
        ],
        "post": [
            "Air India customer satisfaction improvements post
privatization 2022..2024",
            "Enhancements in Air India's in-flight experience post
```

```
privatization"
            ]
        }),
    # Digital Transformation
    ("Digital Transformation",
     {
        "pre": [
            "Air India digital transformation efforts before 2022",
            "Limitations in Air India's booking systems before
privatization",
            "Use of technology in Air India operations before 2022"
        ],
        "post": [
            "Air India digital initiatives post privatization
2022..2024",
            "How has Air India adopted AI and automation since 2022?"
        ]
     }),
    # Fleet Modernization
    ("Fleet Modernization",
     {
        "pre": [
            "Air India fleet challenges before 2022",
            "Aging aircraft in Air India's fleet before
privatization",
            "How did Air India’s fleet modernization efforts fare
before 2022?"
        ],
        "post": [
            "Air India fleet modernization post privatization
2022..2024",
            "New aircraft acquisitions by Air India under Tata Group
2022..2024"
        ]
     }),
    # Sustainability
    ("Sustainability",
     {
        "pre": [
            "Air India sustainability initiatives before 2022",
            "Efforts to reduce emissions by Air India before
privatization",
            "How did Air India address environmental concerns before
2022?"
        ],
        "post": [
            "Air India sustainability programs post privatization
2022..2024",
            "Green technology adoption by Air India under Tata Group
```

```
2022..2024"
        ]
     })
]

# Function to fetch articles for queries
def fetch_articles(query_list, api_key, max_results_per_query=10):
    articles = []
    for query in query_list:
        print(f"Fetching articles for query: {query}")
        search = GoogleSearch({
            "q": query,
            "engine": "google_news",
            "api_key": api_key,
            "sort_by": "date",
        })
        response = search.get_dict()
        if "news_results" in response:
            for result in response["news_results"]
[:max_results_per_query]:
                articles.append({
                    "title": result.get("title"),
                    "link": result.get("link"),
                    "date": result.get("date"),
                    "query": query
                })
    return articles

# Initialize storage for articles
all_articles = []
pre_articles_count = 0
post_articles_count = 0
target_pre_articles = 75  # Ensure at least 75 pre-privatization
articles
target_post_articles = 75  # Ensure at least 75 post-privatization
articles

# Fetch articles for each aspect
for aspect, queries in queries_and_topics:
    print(f"\nFetching articles for aspect: {aspect}")
    # Fetch pre-privatization articles
    if pre_articles_count < target_pre_articles:
        pre_articles = fetch_articles(queries["pre"], api_key,
max_results_per_query=(target_pre_articles - pre_articles_count) //
len(queries["pre"]))
        pre_articles_count += len(pre_articles)
        for article in pre_articles:
            article["aspect"] = aspect
            article["category"] = "pre"
        all_articles.extend(pre_articles)
```

```
    # Fetch post-privatization articles
    if post_articles_count < target_post_articles:
        post_articles = fetch_articles(queries["post"], api_key,
max_results_per_query=(target_post_articles - post_articles_count) //
len(queries["post"]))
        post_articles_count += len(post_articles)
        for article in post_articles:
            article["aspect"] = aspect
            article["category"] = "post"
        all_articles.extend(post_articles)

# Convert collected articles to a DataFrame
df = pd.DataFrame(all_articles)

# Extract publication year from date
df["Publication Year"] = df["date"].str.extract(r"(\d{4})")

# Save the results to a CSV file
output_file = "air_india_case_study_articles.csv"
df.to_csv(output_file, index=False, encoding="utf-8")

print(f"\nArticles saved to {output_file}")
print(f"Pre-Privatization Articles: {pre_articles_count}")
print(f"Post-Privatization Articles: {post_articles_count}")
```

```
Fetching articles for aspect: Operational Efficiency
Fetching articles for query: Air India operational inefficiencies
before 2022
Fetching articles for query: Challenges in Air India's flight delays
and route planning before privatization
Fetching articles for query: How did Air India's operational costs
compare to competitors before 2022?
Fetching articles for query: Air India operational improvements post
privatization 2022..2024
Fetching articles for query: Operational cost reductions at Air India
under Tata Group 2022..2024

Fetching articles for aspect: Market Share
Fetching articles for query: Air India market share growth post
privatization 2022..2024
Fetching articles for query: Market positioning strategies for Air
India under Tata Group 2022..2024

Fetching articles for aspect: Financial Performance
Fetching articles for query: Air India financial turnaround post
privatization 2022..2024
Fetching articles for query: Profitability and cost-cutting measures
at Air India under Tata Group
```

```
Fetching articles for aspect: Customer Experience
Fetching articles for query: Air India customer satisfaction
improvements post privatization 2022..2024
Fetching articles for query: Enhancements in Air India's in-flight
experience post privatization

Fetching articles for aspect: Digital Transformation
Fetching articles for query: Air India digital initiatives post
privatization 2022..2024
Fetching articles for query: How has Air India adopted AI and
automation since 2022?

Fetching articles for aspect: Fleet Modernization
Fetching articles for query: Air India fleet modernization post
privatization 2022..2024
Fetching articles for query: New aircraft acquisitions by Air India
under Tata Group 2022..2024

Fetching articles for aspect: Sustainability
Fetching articles for query: Air India sustainability programs post
privatization 2022..2024
Fetching articles for query: Green technology adoption by Air India
under Tata Group 2022..2024

Articles saved to air_india_case_study_articles.csv
Pre-Privatization Articles: 75
Post-Privatization Articles: 74
```

Web Scrapping: Scrapping content from Articles using BeautifulSoup

```
# Import required libraries
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
import time

# Load the  "air_india_case_study_articles.csv" file
csv_file = "air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Define a function to scrape the article content from the URL
def scrape_article_content(url):
    headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64;
x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.110
Safari/537.3"}

    for _ in range(3):  # Retry up to 3 times
        try:
```

```
            # Send an HTTP GET request
            response = requests.get(url, headers=headers, timeout=10)
            response.raise_for_status()  # Raise HTTPError for bad
responses

            # Parse the HTML content
            soup = BeautifulSoup(response.text, 'html.parser')

            # Extract text from all <p> tags
            paragraphs = soup.find_all('p')
            article_content = ' '.join([para.get_text() for para in
paragraphs])

            return article_content.strip()
        except requests.exceptions.RequestException as e:
            print(f"Error scraping {url}: {e}. Retrying...")
            time.sleep(5)

    # Return None if all retries fail
    print(f"Failed to scrape {url} after 3 attempts.")
    return None

# Add a new 'Content' column to the DataFrame, initialized with None
if "Content" not in df.columns:
    df["Content"] = None

# Loop through the DataFrame rows and scrape each article
for index, row in df.iterrows():
    # Extract data from the row
    url = row.get("link", None)

    # Skip if URL is missing or already processed
    if not url:
        print(f"Skipping article {index + 1}: No URL provided.")
        df.at[index, "Content"] = "Failed"
        continue
    if pd.notnull(row.get("Content")) and row.get("Content") !=
"Failed":
        print(f"Skipping article {index + 1}: Content already
exists.")
        continue

    print(f"Scraping article {index + 1}/{len(df)}: {url}")

    # Scrape the content
    content = scrape_article_content(url)

    # Update the 'Content' column based on success or failure
    if content:
        df.at[index, "Content"] = content
```

```
    else:
        df.at[index, "Content"] = "Failed"  # Mark as failed if
scraping is unsuccessful

    # Pause to avoid overwhelming the server
    time.sleep(3)


# Save the updated DataFrame back to the same
"air_india_case_study_articles.csv" file
df.to_csv(csv_file, index=False, encoding="utf-8", escapechar="\\")
print(f"Updated articles saved to {csv_file}")
```

```
Scraping article 1/149: https://airinsight.com/whats-driving-the-
privatization-of-flag-carriers-in-the-indian-subcontinent/
Scraping article 2/149:
https://aviationa2z.com/index.php/2024/11/12/why-flag-carriers-
privatizing-in-indian-subcontinent/
Scraping article 4/149:
https://timesofindia.indiatimes.com/india/expanding-air-indias-
punctuality-woes-leave-flyers-frustrated/articleshow/113562007.cms
Scraping article 5/149: https://www.dw.com/en/india-is-the-tata-
groups-air-india-revamp-working/a-69096165
Scraping article 6/149: http://www.100knots.com/the-digital-
revolution-of-air-india-a-journey-beyond-silos-and-into-ai/
Scraping article 7/149: https://www.deccanherald.com/business/dh-
deciphers-the-troubles-facing-air-india-since-acquisition-by-tata-
3030758
Scraping article 8/149:
https://m.economictimes.com/industry/transportation/airlines-/-
aviation/2-years-after-privatisation-air-indias-turnaround-is-still-
on-the-tarmac/articleshow/109193891.cms
Scraping article 9/149:
https://www.financialexpress.com/business/airlines-aviation-air-india-
vistara-merger-from-codeshare-flights-to-layoffs-how-this-merger-is-
going-to-affect-passengers-and-employees-3570300/
Scraping article 10/149: https://www.hindustantimes.com/business/air-
india-under-the-tatas-has-a-long-way-to-go-to-become-a-maharajah-
again-101704948648799.html
Scraping article 11/149:
https://aviationa2z.com/index.php/2024/06/10/air-india-maintenance-
widebody-lease-retrofits/
Scraping article 12/149:
https://www.businesstoday.in/industry/aviation/story/from-indigo-to-
vistara-why-airlines-continue-to-face-challenges-430159-2024-05-19
Scraping article 13/149:
https://www.aljazeera.com/economy/2024/5/9/after-years-of-decline-air-
india-is-betting-millions-on-a-comeback
Scraping article 14/149:
https://www.newindianexpress.com/business/2024/Sep/15/killing-vistara-
```

```
may-hurt-tatas-in-long-run
Scraping article 15/149: https://yourstory.com/2024/04/partnership-
meant-vistaras-merger-air-india
Scraping article 16/149: https://www.team-bhp.com/news/air-india-real-
world-reviews-after-tata-takeover
Scraping article 17/149: https://www.reuters.com/graphics/INDIA-
AIRLINES/egpbmkyyqvq/
Scraping article 18/149: https://www.indiatoday.in/diu/story/diu-fog-
bad-weather-airlines-cancellations-delays-2490067-2024-01-17
Scraping article 19/149:
https://www.businesstoday.in/industry/aviation/story/turbulence-at-ai-
express-tatas-must-address-brewing-troubles-on-priority-say-experts-
428950-2024-05-09
Scraping article 20/149: https://www.reuters.com/business/aerospace-
defense/air-travel-hit-by-global-cyber-outage-2024-07-19/
Error scraping https://www.reuters.com/business/aerospace-defense/air-
travel-hit-by-global-cyber-outage-2024-07-19/: 401 Client Error: HTTP
Forbidden for url: https://www.reuters.com/business/aerospace-
defense/air-travel-hit-by-global-cyber-outage-2024-07-19/. Retrying...
Error scraping https://www.reuters.com/business/aerospace-defense/air-
travel-hit-by-global-cyber-outage-2024-07-19/: 401 Client Error: HTTP
Forbidden for url: https://www.reuters.com/business/aerospace-
defense/air-travel-hit-by-global-cyber-outage-2024-07-19/. Retrying...
Error scraping https://www.reuters.com/business/aerospace-defense/air-
travel-hit-by-global-cyber-outage-2024-07-19/: 401 Client Error: HTTP
Forbidden for url: https://www.reuters.com/business/aerospace-
defense/air-travel-hit-by-global-cyber-outage-2024-07-19/. Retrying...
Failed to scrape
https://www.reuters.com/business/aerospace-defense/air-travel-hit-by-
global-cyber-outage-2024-07-19/ after 3 attempts.
Scraping article 21/149:
https://aviationa2z.com/index.php/2024/09/30/air-india-new-fatigue-
policy-for-pilots-cabin-crew/
Scraping article 22/149:
https://www.aljazeera.com/news/2021/10/19/tatas-real-challenge-to-
make-air-india-worth-flying
Scraping article 23/149: https://simpleflying.com/air-india-one-year-
anniversary-as-private-airline/
Scraping article 24/149: https://www.fortuneindia.com/opinion/air-
india-turnaround-why-patience-is-running-out/116937
Scraping article 25/149:
https://sundayguardianlive.com/news/overcoming-hurdles-air-india-
emerges-stronger-ever
Scraping article 26/149:
https://aviationa2z.com/index.php/2024/11/12/why-flag-carriers-
privatizing-in-indian-subcontinent/
Scraping article 27/149:
https://centreforaviation.com/analysis/reports/air-india-ceo-outlines-
initiatives-to-fulfil-post-privatisation-growth-ambitions-696893
```

```
Scraping article 28/149:
https://www.aljazeera.com/economy/2024/5/9/after-years-of-decline-air-india-is-betting-millions-on-a-comeback
Scraping article 29/149: http://www.100knots.com/is-tap-air-portugal-lufthansas-next-target/
Scraping article 30/149: https://simpleflying.com/air-india-widebodies-reach-us-growth-potential/
Scraping article 31/149: https://www.dw.com/en/india-is-the-tata-groups-air-india-revamp-working/a-69096165
Scraping article 32/149: https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/
Error scraping https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/: 403 Client Error: Forbidden for url: https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/. Retrying...
Error scraping https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/: 403 Client Error: Forbidden for url: https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/. Retrying...
Error scraping https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/: 403 Client Error: Forbidden for url: https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/. Retrying...
Failed to scrape https://airlinegeeks.com/2024/07/15/sri-lankan-airlines-seeks-to-restructure-after-failed-privatization-attempt/ after 3 attempts.
Scraping article 33/149:
https://www.suchtv.pk/pakistan/general/item/128958-govt-plans-new-bids-as-pia-privatisation-back-on-track.html
Scraping article 34/149:
https://aviationa2z.com/index.php/2024/10/20/air-india-new-codeshare-with-international-carriers/
Scraping article 35/149: https://thediplomat.com/2024/07/why-the-philippines-chose-to-privatize-its-largest-airport/
Scraping article 36/149:
https://m.economictimes.com/industry/transportation/airlines-/-aviation/2-years-after-privatisation-air-indias-turnaround-is-still-on-the-tarmac/articleshow/109193891.cms
Scraping article 37/149: https://onemileatatime.com/news/srilankan-airlines-collapse/
Scraping article 38/149: https://airinsight.com/it-could-be-curtains-for-srilankan-airlines/
Scraping article 39/149: http://www.100knots.com/wizz-air-defies-odds-squeezes-out-a-profit-amid-engine-woes/
Scraping article 40/149:
```

```
https://www.thehindu.com/business/Industry/air-india-express-cancels-flights-due-to-cabin-crew-shortage/article68152201.ece
Scraping article 41/149:
https://timesofindia.indiatimes.com/business/india-business/air-india-issues-retirees-much-awaited-new-travel-policy/articleshow/107515759.cms
Scraping article 42/149:
https://www.strategy-business.com/article/Wings-of-change-at-Air-India
Scraping article 43/149: https://aeroxplorer.com/articles/air-india-to-reportedly-lease-six-boeing-777s.php
Scraping article 44/149:
https://aviationa2z.com/index.php/2024/05/19/air-india-cuts-kozhikode-while-saudi-airlines-adds-new-flights/
Scraping article 45/149:
https://aviationa2z.com/index.php/2024/06/10/air-india-maintenance-widebody-lease-retrofits/
Scraping article 46/149: https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/
Error scraping https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/: 401 Client Error: HTTP Forbidden for url:
https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/. Retrying...
Error scraping https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/: 401 Client Error: HTTP Forbidden for url:
https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/. Retrying...
Error scraping https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/: 401 Client Error: HTTP Forbidden for url:
https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/. Retrying...
Failed to scrape https://www.reuters.com/world/india/indias-modi-departs-privatisation-plans-will-overhaul-state-run-firms-sources-2024-07-12/ after 3 attempts.
Scraping article 47/149: https://www.mcgill.ca/iasl/article/air-indias-mega-order-more-just-metal
Scraping article 48/149: https://simpleflying.com/air-india-one-year-anniversary-as-private-airline/
Scraping article 49/149: https://onemileatatime.com/news/air-india-upgrades-mumbai-usa-flights/
Scraping article 50/149:
https://www.economist.com/asia/2023/11/23/india-is-seeing-a-massive-aviation-boom
Scraping article 51/149:
https://www2.deloitte.com/us/en/insights/industry/aerospace-defense/aerospace-and-defense-industry-outlook.html
```

```
Scraping article 52/149: https://skift.com/2024/09/17/air-india-vistara-merger-strategic-win-or-tatas-gamble/
Scraping article 53/149: https://in.benzinga.com/content/32500975/what-is-air-indias-share-price-and-how-to-buy
Scraping article 54/149: https://www.statista.com/statistics/588028/passengers-boarded-by-type-by-indian-air-carriers/
Scraping article 55/149: https://simpleflying.com/why-indian-airlines-struggle-to-be-successful/
Scraping article 56/149: https://www.livemint.com/market/stock-market-news/western-carriers-ipo-opens-tomorrow-10-key-points-to-know-before-you-subscribe-to-rs-493-crore-issue-11726114681732.html
Scraping article 57/149: https://www.reuters.com/graphics/INDIA-AIRLINES/egpbmkyyqvq/
Scraping article 58/149: https://www.thehindubusinessline.com/blexplainer/air-indias-game-changer-the-airbus-a350s-revolution-in-indian-aviation/article67672038.ece
Scraping article 59/149: https://simpleflying.com/saf-cost-competitive-jet-fuel/
Scraping article 60/149: https://www.gminsights.com/blogs/top-challenges-of-aviation-industry
Scraping article 61/149: https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/taking-stock-of-the-pandemics-impact-on-global-aviation
Scraping article 62/149: https://www.cbre.com/insights/reports/global-data-center-trends-2023
Error scraping https://www.cbre.com/insights/reports/global-data-center-trends-2023: 403 Client Error: Forbidden for url: https://www.cbre.com/insights/reports/global-data-center-trends-2023. Retrying...
Error scraping https://www.cbre.com/insights/reports/global-data-center-trends-2023: 403 Client Error: Forbidden for url: https://www.cbre.com/insights/reports/global-data-center-trends-2023. Retrying...
Error scraping https://www.cbre.com/insights/reports/global-data-center-trends-2023: 403 Client Error: Forbidden for url: https://www.cbre.com/insights/reports/global-data-center-trends-2023. Retrying...
Failed to scrape https://www.cbre.com/insights/reports/global-data-center-trends-2023 after 3 attempts.
Scraping article 63/149: https://businessmodelanalyst.com/southwest-airlines-swot-analysis/
Scraping article 64/149: https://www.oag.com/blog/what-next-low-cost-carriers-aviation
Scraping article 65/149: https://www2.deloitte.com/us/en/insights/industry/aerospace-defense/advanced-air-mobility-evtol-aircraft.html
```

```
Scraping article 66/149:
https://www.livemint.com/technology/gadgets/macbook-air-and-the-top-3-
alternatives-that-can-match-its-smooth-functionality-and-come-at-a-
more-cost-effective-price-11715847560309.html
Scraping article 67/149: https://www.mckinsey.com/industries/travel-
logistics-and-infrastructure/our-insights/the-six-secrets-of-
profitable-airlines
Scraping article 68/149: https://skift.com/2023/10/17/skift-india-
report-strong-recovery-and-reduced-losses-for-aviation-industry/
Scraping article 69/149:
https://www2.deloitte.com/us/en/insights/industry/oil-and-gas/oil-and-
gas-industry-outlook.html
Scraping article 70/149: https://www.livemint.com/companies/why-air-
india-express-max-8-could-break-indigo-s-stranglehold-on-india-skies-
11703495505131.html
Scraping article 71/149:
https://www.investopedia.com/ask/answers/120314/who-are-teslas-tsla-
main-competitors.asp
Scraping article 72/149: https://www.iea.org/reports/global-ev-
outlook-2023/trends-in-electric-light-duty-vehicles
Scraping article 73/149: https://airlineweekly.skift.com/2023/04/air-
india-mum-on-cost-savings-as-restructuring-progresses/
Scraping article 74/149:
https://www2.deloitte.com/us/en/insights/industry/manufacturing/global
-supply-chain-resilience-amid-disruptions.html
Scraping article 75/149:
https://www.mckinsey.com/industries/financial-services/our-insights/
fintechs-a-new-paradigm-of-growth
Scraping article 76/149:
https://aviationa2z.com/index.php/2024/11/12/why-flag-carriers-
privatizing-in-indian-subcontinent/
Scraping article 77/149:
https://centreforaviation.com/analysis/reports/air-india-ceo-outlines-
initiatives-to-fulfil-post-privatisation-growth-ambitions-696893
Scraping article 78/149:
https://www.aljazeera.com/economy/2024/5/9/after-years-of-decline-air-
india-is-betting-millions-on-a-comeback
Scraping article 79/149: https://simpleflying.com/air-india-vistara-
distinct-airline-codes-adjust-routes/
Scraping article 80/149: https://www.dw.com/en/india-is-the-tata-
groups-air-india-revamp-working/a-69096165
Scraping article 81/149:
https://www.newsbytesapp.com/news/business/air-india-and-vistara-s-
merger-is-now-complete/story
Scraping article 82/149:
https://www.travelandtourworld.com/news/article/air-india-expands-in-
southeast-asia-with-new-delhi-kuala-lumpur-route-appoints-oscar-
travel-services-as-gsa-in-malaysia/
Scraping article 83/149: https://www.businessinsider.com/air-india-
```

```
new-airbus-a350-big-improvement-tour-review-2024-1
Scraping article 84/149:
https://m.economictimes.com/industry/transportation/airlines-/-
aviation/2-years-after-privatisation-air-indias-turnaround-is-still-
on-the-tarmac/articleshow/109193891.cms
Scraping article 85/149: https://skift.com/2023/12/12/air-indias-2024-
makeover-new-uniforms-new-lounges/
Scraping article 86/149: https://timesofindia.indiatimes.com/india/no-
writ-petition-can-be-filed-against-air-india-rules-sc/articleshow/
110187562.cms
Scraping article 87/149: https://www.fortuneindia.com/enterprise/air-
india-two-years-after/115880
Scraping article 88/149:
https://www.strategy-business.com/article/Wings-of-change-at-Air-India
Scraping article 89/149: https://trak.in/stories/air-india-losses-
reduced-by-73-as-it-earns-rs-100-crore-day-becomes-ebitdar-positive/
Scraping article 90/149:
https://www.newindianexpress.com/business/2024/May/26/hits-and-misses-
of-aviation-sector-in-10-yrs
Scraping article 91/149:
https://www.businesstoday.in/union-budget/story/privatisation-plan-
dropped-budget-2024-could-see-a-big-revamp-for-over-200-state-run-
firms-436949-2024-07-12
Scraping article 92/149: https://www.flightglobal.com/analysis/air-
india-foundations-in-place-after-completion-of-mergers/160707.article
Scraping article 93/149:
https://aviationa2z.com/index.php/2024/05/23/adani-group-eyes-25-
private-airports/
Scraping article 94/149: https://simpleflying.com/air-india-one-year-
anniversary-as-private-airline/
Scraping article 95/149: https://www.reuters.com/graphics/INDIA-
AIRLINES/egpbmkyyqvq/
Scraping article 96/149: http://www.100knots.com/is-tap-air-portugal-
lufthansas-next-target/
Scraping article 97/149:
https://aviationa2z.com/index.php/2024/05/12/air-india-adds-12-new-
international-destinations-since-privatization/
Scraping article 98/149:
https://www.forbes.com/sites/jonathanburgos/2024/01/31/singapore-
airlines-bets-big-on-india/
Scraping article 99/149: https://simpleflying.com/air-india-
scrutinized-long-haul-san-francisco-flights-delays/
Scraping article 100/149: https://onemileatatime.com/news/air-india-
a350-routes/
Scraping article 101/149:
https://aviationa2z.com/index.php/2023/07/27/air-india-intends-
flights-to-boston-and-los-angeles/
Scraping article 102/149: https://aviationweek.com/mro/services-
support/fast-5-post-air-india-privatization-aiesl-eyes-new-mro-markets
```

```
Scraping article 103/149: http://www.100knots.com/the-high-stakes-race-to-revive-pia-who-will-take-over/
Scraping article 104/149: https://airlineweekly.skift.com/2023/04/air-india-mum-on-cost-savings-as-restructuring-progresses/
Scraping article 105/149: https://onemileatatime.com/news/air-india-upgrades-mumbai-usa-flights/
Scraping article 106/149: https://simpleflying.com/air-india-update/
Scraping article 107/149: https://simpleflying.com/air-india-north-america-operations-analysis/
Scraping article 108/149: https://www.itln.in/aviation/air-indias-comeback-elevating-presence-in-indian-aviation-and-cargo-1350711
Scraping article 109/149: https://simpleflying.com/air-india-group-consolidated-annual-profit/
Scraping article 110/149: https://www.travelandtourworld.com/news/article/air-india-expands-its-global-network-with-new-flight-to-london-gatwick-from-bengaluru-this-august/
Scraping article 111/149: https://runwaygirlnetwork.com/2022/06/cabin-interior-upgrades-indian-carriers-address-issues/
Scraping article 112/149: https://simpleflying.com/the-sun-sets-on-airasia-india-brand/
Scraping article 113/149: https://aviationa2z.com/index.php/2024/11/11/air-india-vistara-merger-help-tata-500-crore-yearly/
Scraping article 114/149: http://www.100knots.com/klm-cutting-costs-delaying-investments-to-boost-profit-amid-industry-woes/
Scraping article 115/149: https://aviation.direct/en/Air-India-must-save-money%2C-employees-should-share-hotel-rooms
Scraping article 116/149: https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html
Error scraping https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html: 403 Client Error: Forbidden for url: https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html. Retrying...
Error scraping https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html: 403 Client Error: Forbidden for url: https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html. Retrying...
Error scraping https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-124071000158_1.html: 403 Client Error: Forbidden for url: https://www.business-standard.com/companies/news/air-india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-
```

```
124071000158_1.html. Retrying...
Failed to scrape https://www.business-standard.com/companies/news/air-
india-offers-voluntary-separation-as-vistara-merger-impacts-600-jobs-
124071000158_1.html after 3 attempts.
Scraping article 117/149:
https://indianexpress.com/article/business/aviation/icra-indian-
airlines-pricing-power-stable-cost-environment-9551831/
Scraping article 118/149:
https://www.indiatoday.in/business/story/air-india-vistara-merger-
likely-to-be-finalised-on-november-12-2590580-2024-08-30
Scraping article 119/149:
https://english.mathrubhumi.com/news/money/air-india-cuts-losses-
under-tata-1.9881234
Scraping article 120/149:
https://www.newindianexpress.com/business/2024/Sep/20/air-india-hired-
9000-employees-in-two-years-campbell-wilson
Scraping article 121/149: http://www.100knots.com/cebu-pacifics-bold-
expansion-70-airbus-a321neo-jets-affirmed-in-major-fleet-deal/
Scraping article 122/149:
https://m.economictimes.com/industry/transportation/airlines-/-
aviation/2-years-after-privatisation-air-indias-turnaround-is-still-
on-the-tarmac/articleshow/109193891.cms
Scraping article 123/149: http://www.100knots.com/fatigue-or-fear-air-
indias-new-policy-sparks-debate-on-pilot-reporting/
Scraping article 124/149:
https://www.business-standard.com/companies/news/2-500-job-cuts-
inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html
Error scraping https://www.business-standard.com/companies/news/2-500-
job-cuts-inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html: 403 Client Error: Forbidden for url:
https://www.business-standard.com/companies/news/2-500-job-cuts-
inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html. Retrying...
Error scraping https://www.business-standard.com/companies/news/2-500-
job-cuts-inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html: 403 Client Error: Forbidden for url:
https://www.business-standard.com/companies/news/2-500-job-cuts-
inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html. Retrying...
Error scraping https://www.business-standard.com/companies/news/2-500-
job-cuts-inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html: 403 Client Error: Forbidden for url:
https://www.business-standard.com/companies/news/2-500-job-cuts-
inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html. Retrying...
Failed to scrape https://www.business-standard.com/companies/news/2-
500-job-cuts-inevitable-in-tata-steel-s-uk-operations-ceo-narendran-
124060200157_1.html after 3 attempts.
```

```
Scraping article 125/149: http://www.100knots.com/is-air-indias-new-cabin-crew-policy-a-game-changer-or-a-cause-for-concern/
Scraping article 126/149: http://www.100knots.com/tata-groups-journey-to-make-air-india-world-class-amidst-fleet-refits-and-global-challenges/
Scraping article 127/149: http://www.100knots.com/air-indias-bold-takeoff-aiming-for-global-dominance-again/
Scraping article 128/149: http://www.100knots.com/vietnam-airlines-bold-plans-for-2024/
Scraping article 129/149: http://www.100knots.com/air-india-vistara-merger-navigating-continuity-and-transformation-in-indias-skies/
Scraping article 130/149: http://www.100knots.com/what-happens-next-for-300-air-india-employees-left-out-in-the-cold/
Scraping article 131/149: http://www.100knots.com/how-is-air-india-transforming-under-tata-groups-leadership/
Scraping article 132/149: http://www.100knots.com/fly91-set-to-take-flight-in-february-2024/
Scraping article 133/149: https://trak.in/stories/air-india-losses-reduced-by-73-as-it-earns-rs-100-crore-day-becomes-ebitdar-positive/
Scraping article 134/149: https://urbanacres.in/tata-owned-air-india-eyes-loss-reduction-in-fy24-after-substantial-fy23-setback/
Scraping article 135/149: https://upstox.com/news/business-news/latest-updates/air-india-hired-9-000-employees-in-two-years-says-air-india-chief/article-118058/
Scraping article 136/149: https://www.livemint.com/companies/people/vistara-ceo-kannan-hopes-for-cost-reduction-in-the-budget-11704731274584.html
Scraping article 137/149: https://www.newindianexpress.com/business/2024/May/26/hits-and-misses-of-aviation-sector-in-10-yrs
Scraping article 138/149: https://timesofindia.indiatimes.com/business/india-business/hundreds-of-flights-to-be-cancelled-till-march-will-mumbai-airport-flight-restrictions-lead-to-rising-airfares/articleshow/107678228.cms
Scraping article 139/149: https://www.flightglobal.com/analysis/air-india-foundations-in-place-after-completion-of-mergers/160707.article
Scraping article 140/149: https://aviationa2z.com/index.php/2024/11/12/why-flag-carriers-privatizing-in-indian-subcontinent/
Scraping article 141/149: https://www.reuters.com/graphics/INDIA-AIRLINES/egpbmkyyqvq/
Scraping article 142/149: http://www.100knots.com/aix-to-say-goodbye-to-business-class-the-big-change-coming-in-2025/
Scraping article 143/149: https://aviationa2z.com/index.php/2024/10/12/air-india-begins-a320-aircraft-transfer-to-ai-express/
Scraping article 144/149: https://airlineweekly.skift.com/2023/04/air-india-mum-on-cost-savings-as-restructuring-progresses/
Scraping article 145/149:
```

```
https://www.hindustantimes.com/business/singapore-airlines-to-invest-rs-3-194-5-crore-more-into-air-india-after-vistara-merger-for-25-1-stake-101731292945808.html
Scraping article 146/149: http://www.100knots.com/lufthansa-eyes-potential-stake-in-airbaltic-a-strategic-move-in-european-airline-consolidation/
Scraping article 147/149: https://aviationa2z.com/index.php/2024/05/19/air-india-cuts-kozhikode-while-saudi-airlines-adds-new-flights/
Scraping article 148/149: https://indianexpress.com/article/business/aviation/air-india-vistara-merger-aircraft-line-maintenance-dgca-9504544/
Scraping article 149/149: https://www.livemint.com/companies/news/vistara-crisis-from-roster-revision-to-slashing-daily-operations-airline-ceo-unveils-plan-to-address-pilot-shortage-11712196159558.html
Updated articles saved to air_india_case_study_articles.csv
```

Data Preprocessing: Data Cleaning

```python
import pandas as pd

# Load the updated CSV file with the scraped content
csv_file = "air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Remove rows where the 'Content' column is "Failed"
df_cleaned = df[df["Content"] != "Failed"].copy()

# Save the cleaned DataFrame back to the same CSV file
df_cleaned.to_csv(csv_file, index=False, encoding="utf-8")

print(f"Removed rows with 'Failed' content. \nUpdated file saved to {csv_file}.")
```

```
Removed rows with 'Failed' content.
Updated file saved to air_india_case_study_articles.csv.
```

```python
# Display the shape of the dataset
print(f"Dataset Shape: {df.shape}")
print(f"Number of Rows: {df.shape[0]}, Number of Columns: {df.shape[1]}")

# Display the column names
print("\nColumn Names:")
print(df.columns.tolist())

# Preview the first 5 rows of the dataset
print("\nDataset Preview:")
print(df.head())
```

```python
# Summary of missing values
print("\nMissing Values Summary:")
print(df.isnull().sum())
```

```
Dataset Shape: (149, 8)
Number of Rows: 149, Number of Columns: 8

Column Names:
['title', 'link', 'date', 'query', 'aspect', 'category', 'Publication
Year', 'Content']

Dataset Preview:
                                                title  \
0  What’s Driving the Privatization of Flag Carri...
1  Why Flag Carriers Are Privatizing in the India...
2  Air India-Vistara Merger: Strategic Win or Tat...
3  Expanding Air India's punctuality woes leave f...
4  India: Is the Tata Group's Air India revamp wo...

                                                 link  \
0  https://airinsight.com/whats-driving-the-priva...
1  https://aviationa2z.com/index.php/2024/11/12/w...
2  https://skift.com/2024/09/17/air-india-vistara...
3  https://timesofindia.indiatimes.com/india/expa...
4  https://www.dw.com/en/india-is-the-tata-groups...

                              date  \
0  11/19/2024, 10:55 PM, +0000 UTC
1  11/12/2024, 05:44 PM, +0000 UTC
2  09/17/2024, 07:00 AM, +0000 UTC
3  09/22/2024, 07:00 AM, +0000 UTC
4  05/16/2024, 07:00 AM, +0000 UTC

                                               query
aspect  \
0  Air India operational inefficiencies before 2022  Operational
Efficiency
1  Air India operational inefficiencies before 2022  Operational
Efficiency
2  Air India operational inefficiencies before 2022  Operational
Efficiency
3  Air India operational inefficiencies before 2022  Operational
Efficiency
4  Air India operational inefficiencies before 2022  Operational
Efficiency

  category  Publication Year  \
0      pre              2024
1      pre              2024
```

```
2      pre              2024
3      pre              2024
4      pre              2024

                                                Content
0  Commercial Aviation News and Analysis  \n     ...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

Missing Values Summary:
title               0
link                0
date                0
query               0
aspect              0
category            0
Publication Year    0
Content             0
dtype: int64
```

```python
import pandas as pd
import numpy as np

# Load the "air_india_case_study_articles.csv" dataset
csv_file = "air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Remove duplicates
print("Removing duplicate rows...")
df.drop_duplicates(inplace=True)

# Handle missing values
print("Handling missing values...")
df = df[df["Content"].notnull() & (df["Content"] != "Failed")].copy()
df["date"] = df["date"].fillna("Unknown Date")

# Extract and clean the 'Publication Year' column
print("Extracting publication year from date...")
def parse_year(date):
    try:
        return pd.to_datetime(date, errors="coerce").year
    except:
        return 0

df["Publication Year"] = df["date"].apply(parse_year)

# Clean and preprocess the 'Content' column
print("Cleaning article content...")
```

```
df["Content"] = df["Content"].str.replace(r"\s+", " ",
regex=True).str.strip()

# Add a 'Word Count' column
print("Adding word count for article content...")
df["Word Count"] = df["Content"].apply(lambda x: len(x.split()) if
pd.notnull(x) else 0)

# Filter articles with insufficient word count (less than 50 words)
print("Filtering articles with fewer than 50 words...")
df = df.loc[df["Word Count"] >= 50].copy()

# Classify articles into aspects based on the 'query' column
print("Classifying articles into aspects...")
aspect_mapping = {
    "operational efficiency": "Operational Efficiency",
    "market share": "Market Share",
    "financial performance": "Financial Performance",
    "customer experience": "Customer Satisfaction",
    "fleet modernization": "Fleet Modernization",
    "leadership": "Leadership",
    "digital transformation": "Technology & Innovation",
    "route expansion": "Route Network Expansion",
    "profitability": "Financial Performance",
    "sustainability": "Sustainability",
    "csr": "Corporate Social Responsibility",
}
df["Aspect"] = df["query"].apply(
    lambda q: next((aspect for key, aspect in aspect_mapping.items()
if pd.notnull(q) and key in q.lower()), "Unknown Aspect")
)

# Reset index after cleaning
print("Resetting index after cleaning...")
df.reset_index(drop=True, inplace=True)

# Save the cleaned dataset
output_file = "cleaned_air_india_case_study_articles.csv"
df.to_csv(output_file, index=False, encoding="utf-8")

print(f"Data cleaning complete. Cleaned dataset saved to
{output_file}.")
print(f"Final dataset shape: {df.shape}")

Removing duplicate rows...
Handling missing values...
Extracting publication year from date...
Cleaning article content...
Adding word count for article content...
Filtering articles with fewer than 50 words...
```

```
Classifying articles into aspects...
Resetting index after cleaning...
Data cleaning complete. Cleaned dataset saved to
cleaned_air_india_case_study_articles.csv.
Final dataset shape: (142, 10)
```

```
# Load both datasets
df_original = pd.read_csv("air_india_case_study_articles.csv")
df_cleaned = pd.read_csv("cleaned_air_india_case_study_articles.csv")

# Merge the 'Publication Year' column from the original dataset
df_cleaned["Publication Year"] = df_original["Publication Year"]

# Save the updated cleaned dataset
df_cleaned.to_csv("cleaned_air_india_case_study_articles.csv",
index=False, encoding="utf-8")
print("Recovered 'Publication Year' and saved updated file.")
```

```
Recovered 'Publication Year' and saved updated file.
```

```
# Import required libraries
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer

# Download necessary NLTK resources
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

# Load the cleaned dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Initialize stopwords and lemmatizer
stop_words = set(stopwords.words("english"))
lemmatizer = WordNetLemmatizer()

# Define a function to preprocess text
def preprocess_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove punctuation
    text = re.sub(r'[^\w\s]', '', text)
    # Tokenize using regex
    tokens = re.findall(r'\b\w+\b', text)  # Split into words
    # Remove stopwords and lemmatize
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word
not in stop_words]
```

```
    # Join tokens back into a single string
    return ' '.join(tokens)

# Apply the alternative tokenization method
df["Processed Content"] = df["Content"].apply(preprocess_text)

# Save the updated dataset back to the same file
df.to_csv(csv_file, index=False, encoding="utf-8")

print(f"Text preprocessing complete. Updated dataset saved back to
{csv_file}.")
print(f"Final dataset shape: {df.shape}")
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!

Text preprocessing complete. Updated dataset saved back to
cleaned_air_india_case_study_articles.csv.
Final dataset shape: (142, 11)
```

Exploratory Data Analysis: Numerical Analysis

```
# Import the pandas library
import pandas as pd

# Load the "cleaned_air_india_case_study_articles.csv" dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Get Dataset Overview
print("Dataset Overview:")
print(f"Shape: {df.shape}")
print(df.info())

print("\nSummary Statistics:")
print(df.describe())

# Get Word Count Statistics
print("\nWord Count Statistics:")
print(f"Average Word Count: {df['Word Count'].mean():.2f}")
print(f"Min Word Count: {df['Word Count'].min()}")
print(f"Max Word Count: {df['Word Count'].max()}")

# Get Distribution of Articles by Aspect
aspect_counts = df['Aspect'].value_counts()
print("\nNumber of Articles by Aspect:")
```

```python
print(aspect_counts)

# Get Pre- and Post-Privatization Article Counts based on 'category' column
if "category" in df.columns:
    pre_priv = df[df["category"] == "pre"].shape[0]
    post_priv = df[df["category"] == "post"].shape[0]
    print(f"\nPre-Privatization Articles: {pre_priv}")
    print(f"Post-Privatization Articles: {post_priv}")
else:
    print("\nCategory column is missing.")
```

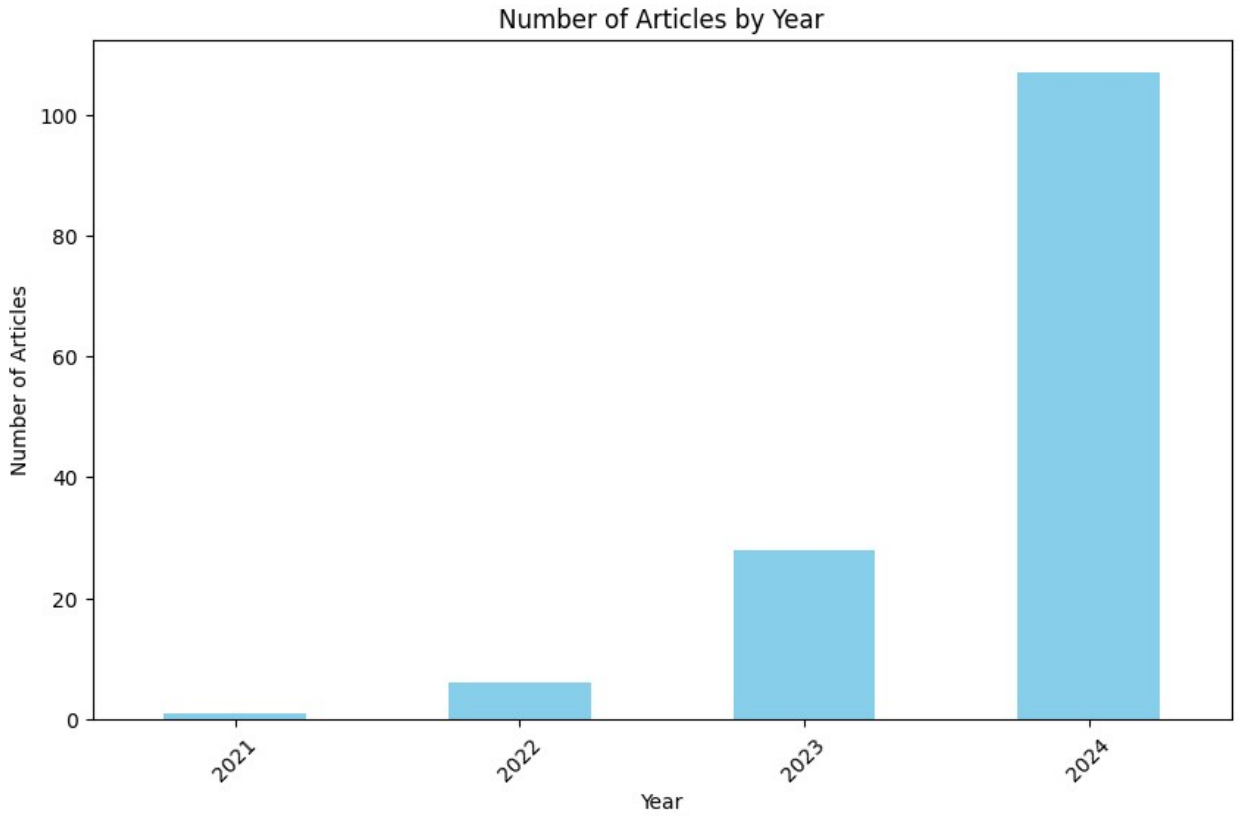
```
Dataset Overview:
Shape: (142, 11)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 142 entries, 0 to 141
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   title              142 non-null    object
 1   link               142 non-null    object
 2   date               142 non-null    object
 3   query              142 non-null    object
 4   aspect             142 non-null    object
 5   category           142 non-null    object
 6   Publication Year   142 non-null    int64 
 7   Content            142 non-null    object
 8   Word Count         142 non-null    int64 
 9   Aspect             142 non-null    object
 10  Processed Content  142 non-null    object
dtypes: int64(2), object(9)
memory usage: 12.3+ KB
None

Summary Statistics:
       Publication Year    Word Count
count        142.000000    142.000000
mean        2023.697183   2324.957746
std            0.583332   2954.078224
min         2021.000000     86.000000
25%         2024.000000    691.000000
50%         2024.000000    963.000000
75%         2024.000000   1962.250000
max         2024.000000  16187.000000

Word Count Statistics:
Average Word Count: 2324.96
Min Word Count: 86
Max Word Count: 16187
```

```
Number of Articles by Aspect:
Aspect
Unknown Aspect    142
Name: count, dtype: int64

Pre-Privatization Articles: 70
Post-Privatization Articles: 72
```

Exploratory Data Analysis: Visualization

Distribution of Articles by Publication Year

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud

# Load the dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Ensure 'Publication Year' is numeric and drop missing values
df["Publication Year"] = pd.to_numeric(df["Publication Year"],
errors="coerce")
df.dropna(subset=["Publication Year"], inplace=True)

# Validate if the 'Publication Year' column is not empty
if not df["Publication Year"].empty:
    # Distribution of Articles by Year
    plt.figure(figsize=(10, 6))
    df['Publication
Year'].value_counts().sort_index().plot(kind='bar', color='skyblue')
    plt.title("Number of Articles by Year")
    plt.xlabel("Year")
    plt.ylabel("Number of Articles")
    plt.xticks(rotation=45)
    plt.show()
else:
    print("No valid data in 'Publication Year' to plot.")
```

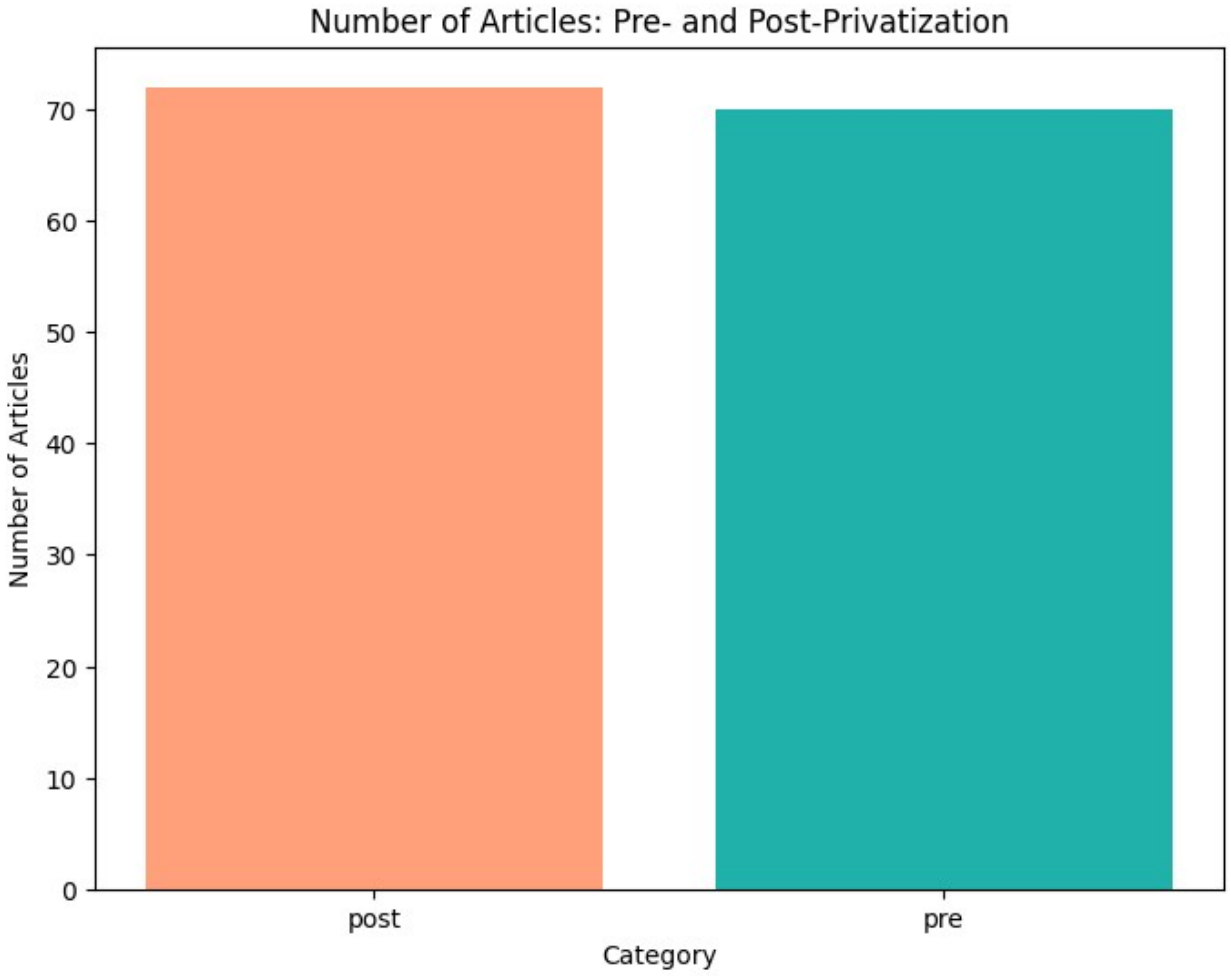


The number of articles significantly increases over the years, peaking in 2024. Minimal activity in earlier years (2021 and 2022).

Number of Articles Categorized by Pre- and Post-Privatization

```
# Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud

# Load the dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Ensure 'Word Count' is numeric
df["Word Count"] = pd.to_numeric(df["Word Count"], errors="coerce")

# Distribution of Articles by Category (Pre and Post)
plt.figure(figsize=(8, 6))
category_counts = df["category"].value_counts()
plt.bar(category_counts.index, category_counts.values,
color=['#FFA07A', '#20B2AA'])
plt.title("Number of Articles: Pre- and Post-Privatization")
plt.xlabel("Category")
```

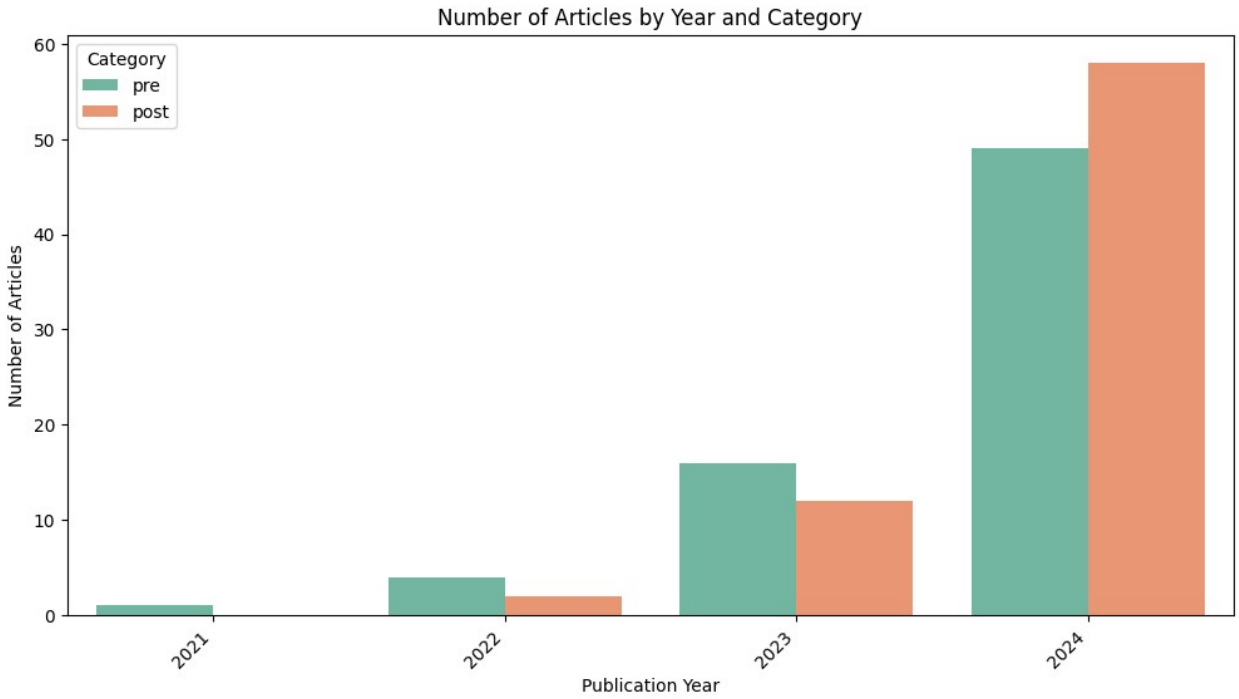
```
plt.ylabel("Number of Articles")
plt.xticks(rotation=0)
plt.show()
```



Almost equal number of articles for both pre and post privatization

Article Counts by Year and Category

```
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x="Publication Year", hue="category",
palette="Set2")
plt.title("Number of Articles by Year and Category")
plt.xlabel("Publication Year")
plt.ylabel("Number of Articles")
plt.xticks(rotation=45, ha="right")
plt.legend(title="Category")
plt.show()
```

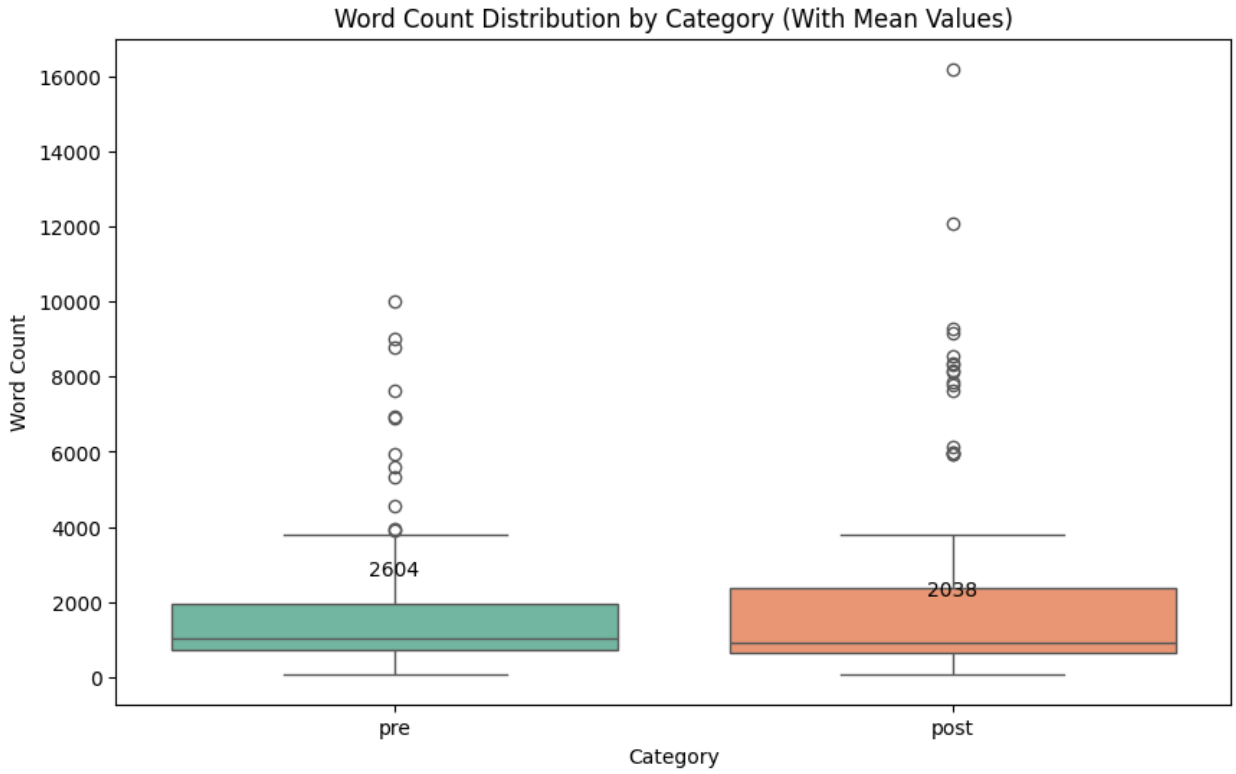


2024 has the highest number of articles for both "pre" and "post" categories. 2021 and 2022 have significantly fewer articles, possibly due to less activity or relevance during those years.

Distribution of Word Counts Across Pre- and Post-Privatization Categories

```
# Calculate Mean Word Count by Category
mean_counts = df.groupby("category")["Word Count"].mean()

# Plot Word Count Distribution by Category with Means
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x="category", y="Word Count", palette="Set2")  #
Use 'df' instead of 'filtered_df'
for i, category in enumerate(mean_counts.index):
    plt.text(i, mean_counts[category], f"{mean_counts[category]:.0f}",
ha="center", va="bottom", fontsize=10)
plt.title("Word Count Distribution by Category (With Mean Values)")
plt.xlabel("Category")
plt.ylabel("Word Count")
plt.show()
```

```
<ipython-input-13-36075759c07b>:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.boxplot(data=df, x="category", y="Word Count", palette="Set2")
# Use 'df' instead of 'filtered_df'
```



Articles in the "pre" category have a slightly higher mean word count (2604) compared to the "post" category (2038). Both categories exhibit outliers with significantly higher word counts.

Word Cloud Visualization for Pre and Post Articles

```
from wordcloud import WordCloud

# Generate word clouds for each category
for category in ["pre", "post"]:
    text = " ".join(df[df["category"] == category]
["Content"].dropna())
    wordcloud = WordCloud(width=800, height=400,
background_color="white").generate(text)

    plt.figure(figsize=(10, 6))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.title(f"Word Cloud for {category.capitalize()} Articles")
    plt.axis("off")
    plt.show()
```

Word Cloud for Pre Articles

Word Cloud for Post Articles

Pre-Privatization: Focused on "airline," "India," "aircraft," "year," and "market". Articles often highlighted challenges and inefficiencies.

Post-Privatization: New terms like "fleet," "Vistara," "Boeing," and "business class" emerged, reflecting operational improvements and strategic growth.

Topic Modeling & Analysis using LDA

```python
# Import necessary libraries
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.preprocessing import normalize
import numpy as np

# Load the cleaned dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Ensure there is no missing data in the 'Processed Content' column
df["Processed Content"] = df["Processed Content"].fillna("")

# Vectorize the text data
vectorizer = CountVectorizer(max_df=0.9, min_df=10,
stop_words="english")
content_matrix = vectorizer.fit_transform(df["Processed Content"])

# Fit LDA to the data
n_topics = 5  # Number of topics
lda = LatentDirichletAllocation(n_components=n_topics,
random_state=42)
lda.fit(content_matrix)

# Extract top words for each topic
words = vectorizer.get_feature_names_out()
n_top_words = 10  # Number of top words per topic

# Store topic keywords and labels
topic_keywords = []
for idx, topic in enumerate(lda.components_):
    top_words = [words[i] for i in topic.argsort()[-n_top_words:]]
    topic_keywords.append(top_words)
    print(f"Topic {idx + 1}: {', '.join(top_words)}")

# Assign labels to topics based on extracted keywords
topic_labels = {
    0: "Domestic Market and Fleet Expansion",
    1: "Financial Performance and Investments",
    2: "Passenger Services and Global Strategy",
    3: "Supply Chain and Sustainability",
    4: "Customer Experience and Operations",
}

# Display named topics
for idx, keywords in enumerate(topic_keywords):
    print(f"Topic {idx + 1} - {topic_labels[idx]}: {',
'.join(keywords)}")

# Assign dominant topics to each article
```

```
topic_distributions = lda.transform(content_matrix)  # Get topic
distribution for each article
dominant_topics = np.argmax(topic_distributions, axis=1)  # Find the
dominant topic index for each article

# Add dominant topic labels to the dataframe
df["Topic"] = dominant_topics
df["Topic Label"] = df["Topic"].apply(lambda x: topic_labels[x])

# Save the dataset with topic labels
output_file = "air_india_analysis_dataset.csv"
df.to_csv(output_file, index=False, encoding="utf-8")

print(f"Dataset with topics saved to {output_file}")
```

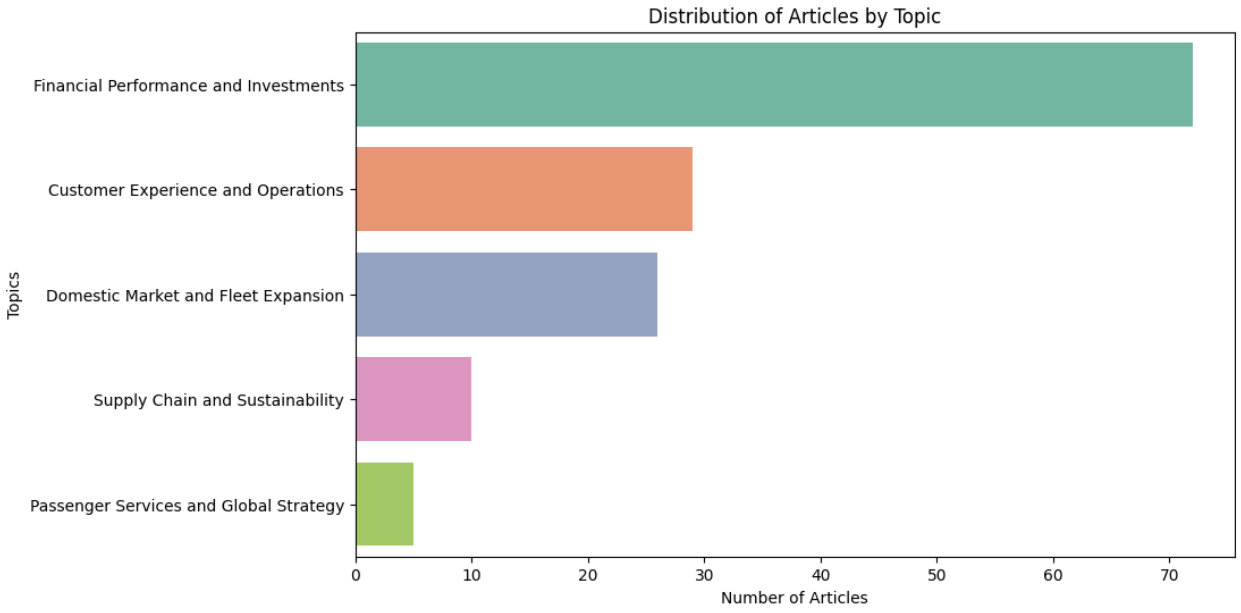
```
Topic 1: boeing, new, significant, service, 2024, industry, aircraft,
aviation, air, airline
Topic 2: group, aviation, new, tata, year, flight, carrier, aircraft,
airline, air
Topic 3: delhi, travel, class, route, airline, business, a350, flight,
new, air
Topic 4: chain, 2024, energy, year, experience, 2023, supply,
technology, industry, company
Topic 5: percent, model, cost, growth, company, year, 2022, car, sale,
market
Topic 1 - Domestic Market and Fleet Expansion: boeing, new,
significant, service, 2024, industry, aircraft, aviation, air, airline
Topic 2 - Financial Performance and Investments: group, aviation, new,
tata, year, flight, carrier, aircraft, airline, air
Topic 3 - Passenger Services and Global Strategy: delhi, travel,
class, route, airline, business, a350, flight, new, air
Topic 4 - Supply Chain and Sustainability: chain, 2024, energy, year,
experience, 2023, supply, technology, industry, company
Topic 5 - Customer Experience and Operations: percent, model, cost,
growth, company, year, 2022, car, sale, market
Dataset with topics saved to air_india_analysis_dataset.csv
```

Distribution of Articles Across Topics

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

# Count the number of articles per topic
topic_counts = df["Topic Label"].value_counts()

# Plot the topic distribution as a bar chart
plt.figure(figsize=(10, 6))
sns.barplot(x=topic_counts.values, y=topic_counts.index,
palette="Set2")
plt.title("Distribution of Articles by Topic")
```

```
plt.xlabel("Number of Articles")
plt.ylabel("Topics")
plt.show()
```

```
<ipython-input-17-78ec8beace1d>:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x=topic_counts.values, y=topic_counts.index,
palette="Set2")
```

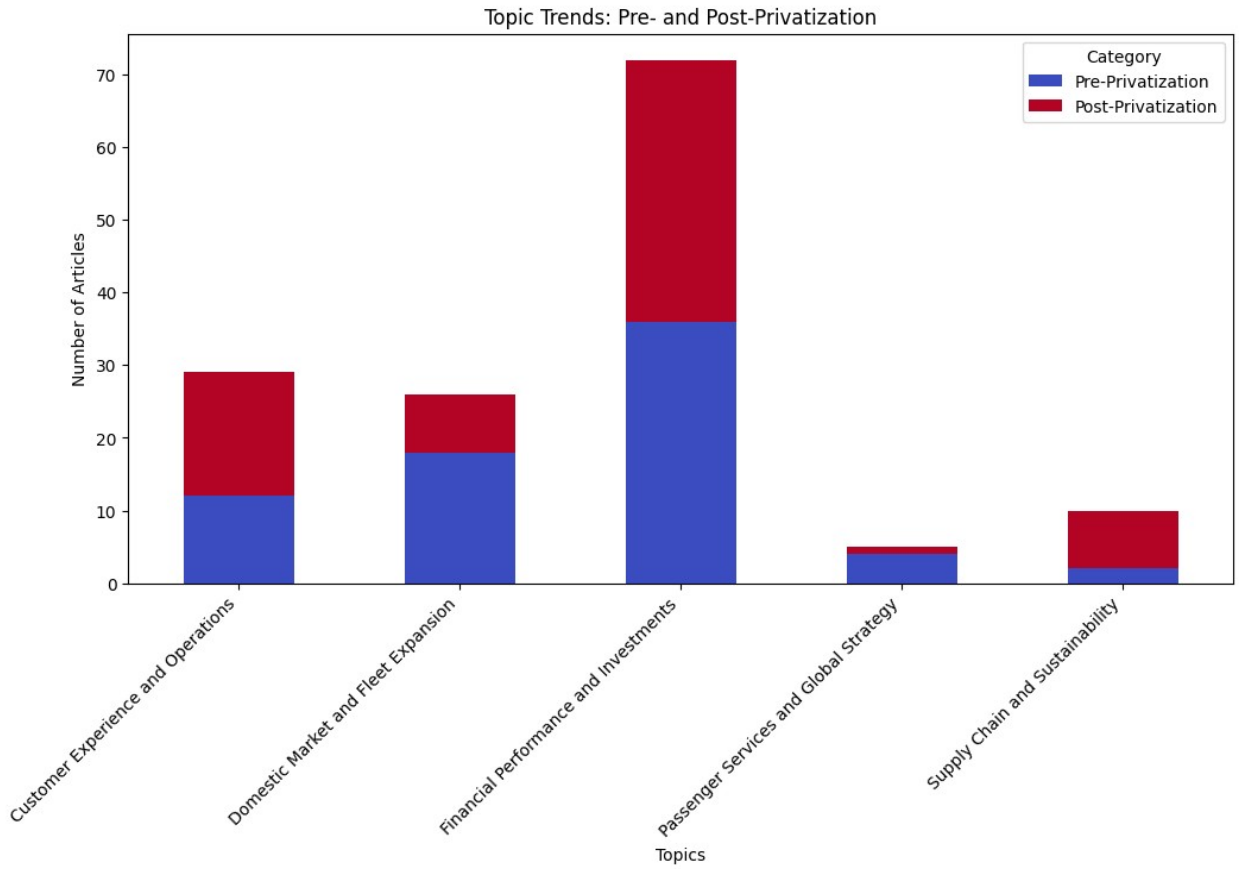


Financial Performance and Investments is the most discussed topic, with over 70 articles. Other significant topics include Customer Experience and Operations and Domestic Market and Fleet Expansion.

Distribution of Articles Across Topics by pre and post privatization

```
# Count topics by category (pre/post)
topic_trend = df.groupby(["Topic Label",
"category"]).size().unstack(fill_value=0)

# Plot the comparison
topic_trend.plot(kind="bar", figsize=(12, 6), stacked=True,
colormap="coolwarm")
plt.title("Topic Trends: Pre- and Post-Privatization")
plt.xlabel("Topics")
plt.ylabel("Number of Articles")
plt.xticks(rotation=45, ha="right")
plt.legend(title="Category", labels=["Pre-Privatization", "Post-
```

```
Privatization"])
plt.show()
```



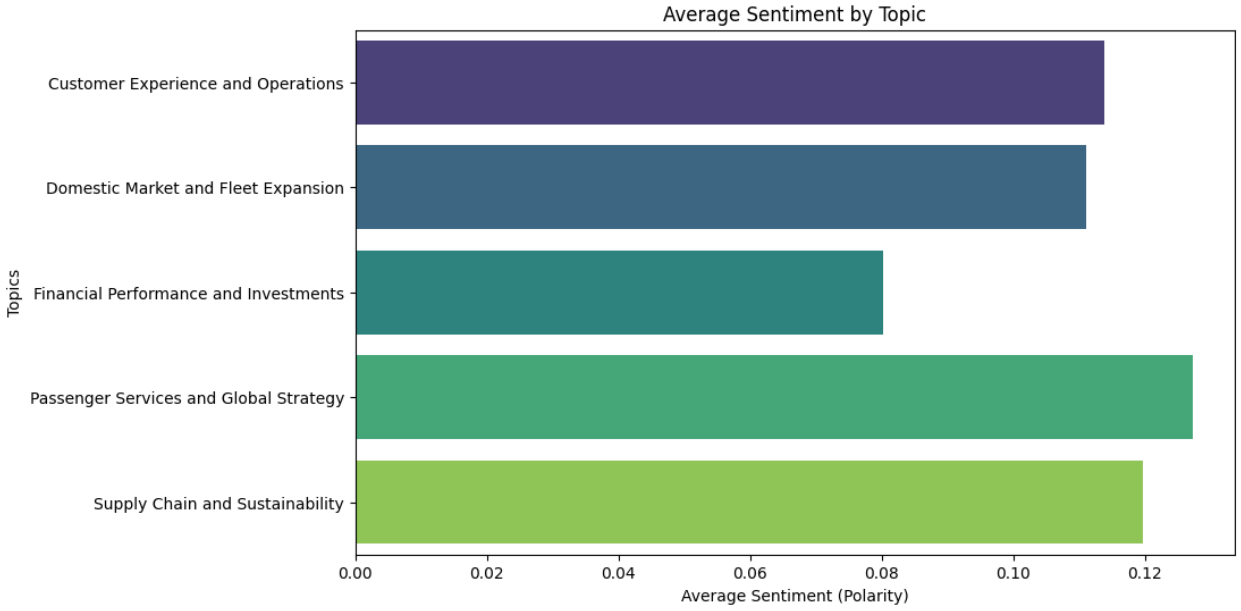
Financial Performance and Investments dominates both periods, especially post-privatization. Topics like Customer Experience and Operations and Domestic Market and Fleet Expansion show balanced representation across both categories.

Average Sentiment by Topic Across Articles

```
from textblob import TextBlob

# Define a function to compute sentiment polarity
def analyze_sentiment(text):
    return TextBlob(text).sentiment.polarity # sentiment.polarity
ranges from -1 (negative) to +1 (positive)

# Apply sentiment analysis to the 'Processed Content' column
df["Sentiment"] = df["Processed Content"].apply(analyze_sentiment)

# Average sentiment per topic
topic_sentiment = df.groupby("Topic Label")["Sentiment"].mean()

# Plot sentiment by topic
```

```
plt.figure(figsize=(10, 6))
sns.barplot(x=topic_sentiment.values, y=topic_sentiment.index,
palette="viridis")
plt.title("Average Sentiment by Topic")
plt.xlabel("Average Sentiment (Polarity)")
plt.ylabel("Topics")
plt.show()
```

```
<ipython-input-19-965404848bc8>:15: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x=topic_sentiment.values, y=topic_sentiment.index,
palette="viridis")
```

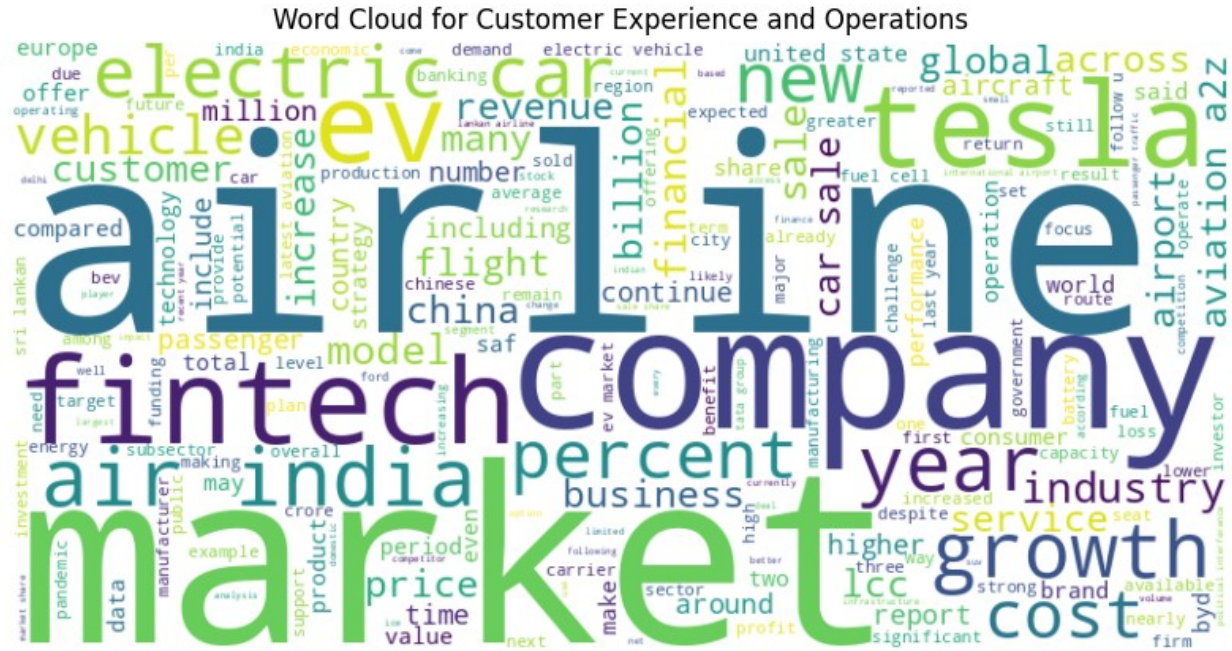


Supply Chain and Sustainability has the highest positive sentiment. Passenger Services and Global Strategy also shows favorable sentiments.

Word Cloud for Customer Experience and Operations

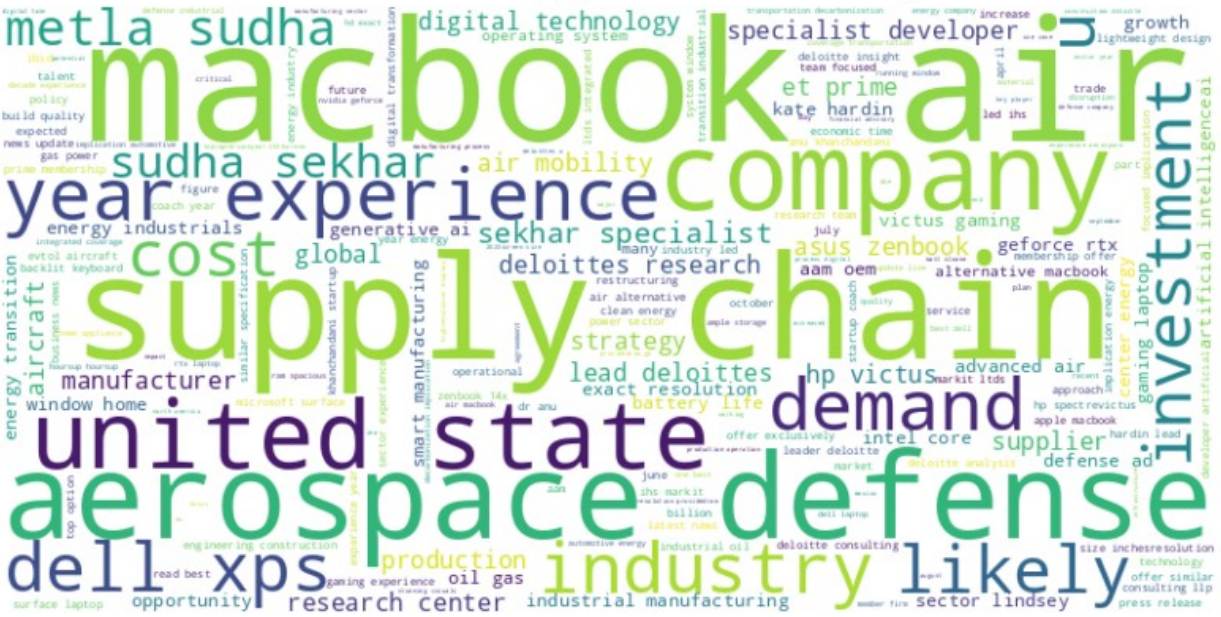
```
from wordcloud import WordCloud

# Generate word clouds for each topic
for topic in df["Topic Label"].unique():
    topic_data = df[df["Topic Label"] == topic]["Processed Content"]
    text = " ".join(topic_data)

    wordcloud = WordCloud(width=800, height=400,
background_color="white").generate(text)
```

```
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.title(f"Word Cloud for {topic}")
    plt.show()
```



Word Cloud for Domestic Market and Fleet Expansion

Word Cloud for Supply Chain and Sustainability



Topic Statistics & Sentiment Analysis

```
# Example of summarizing topic insights
for topic, words in zip(topic_labels.values(), topic_keywords):
    print(f"**{topic}:**")
    print(f"- Key Words: {', '.join(words)}")
    print(f"- Number of Articles: {df[df['Topic Label'] ==
topic].shape[0]}")
    avg_sentiment = df[df["Topic Label"] == topic]["Sentiment"].mean()
    print(f"- Average Sentiment: {avg_sentiment:.2f}\n")
```

```
**Domestic Market and Fleet Expansion:**
- Key Words: boeing, new, significant, service, 2024, industry,
aircraft, aviation, air, airline
- Number of Articles: 26
- Average Sentiment: 0.11

**Financial Performance and Investments:**
- Key Words: group, aviation, new, tata, year, flight, carrier,
aircraft, airline, air
- Number of Articles: 72
- Average Sentiment: 0.08

**Passenger Services and Global Strategy:**
- Key Words: delhi, travel, class, route, airline, business, a350,
flight, new, air
- Number of Articles: 5
- Average Sentiment: 0.13

**Supply Chain and Sustainability:**
```

```
- Key Words: chain, 2024, energy, year, experience, 2023, supply,
technology, industry, company
- Number of Articles: 10
- Average Sentiment: 0.12

**Customer Experience and Operations:**
- Key Words: percent, model, cost, growth, company, year, 2022, car,
sale, market
- Number of Articles: 29
- Average Sentiment: 0.11
```

Article count in every Topic & Sentiment Category

```
# Categorize sentiment based on thresholds
def categorize_sentiment(score):
    if score > 0.05:
        return "Positive"
    elif score < -0.05:
        return "Negative"
    else:
        return "Neutral"

# Calculate sentiment category for each topic
topic_sentiment = df.groupby("Topic Label")["Sentiment"].mean()
sentiment_category = topic_sentiment.apply(categorize_sentiment)

# Combine results
topic_summary = pd.DataFrame({
    "Topic": topic_sentiment.index,
    "Number of Articles": df["Topic Label"].value_counts(),
    "Sentiment Category": sentiment_category
})

# Display the updated summary
print(topic_summary)
```

```
Topic  \
Topic Label

Customer Experience and Operations          Customer Experience and
Operations
Domestic Market and Fleet Expansion        Domestic Market and Fleet
Expansion
Financial Performance and Investments    Financial Performance and
Investments
Passenger Services and Global Strategy  Passenger Services and Global
Strategy
Supply Chain and Sustainability                Supply Chain and
```

```
Sustainability

                                          Number of Articles Sentiment
Category
Topic Label

Customer Experience and Operations                        29
Positive
Domestic Market and Fleet Expansion                       26
Positive
Financial Performance and Investments                     72
Positive
Passenger Services and Global Strategy                     5
Positive
Supply Chain and Sustainability                           10
Positive
```

Article count in every Topic & Sentiment Category with pre & post category count

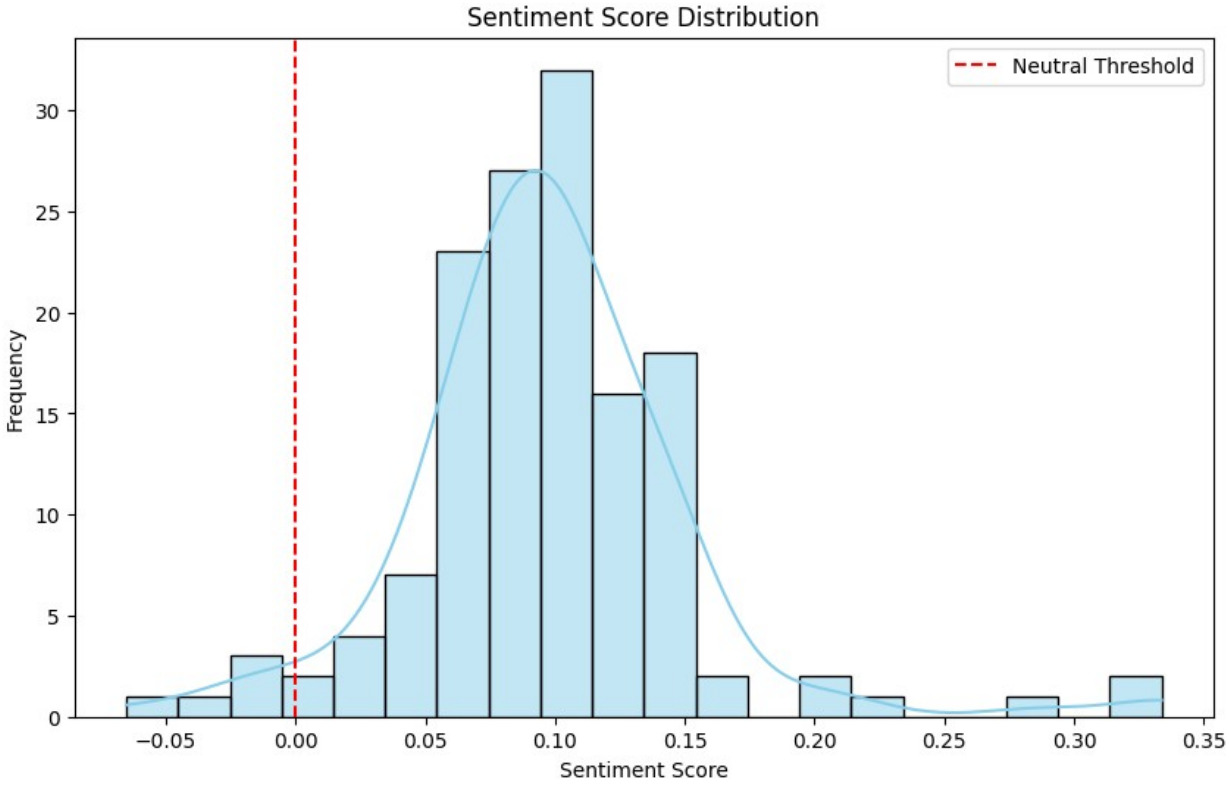
```
# Ensure all topics are present in both pre- and post-privatization
all_topics = df["Topic Label"].unique()

# Split the data into pre- and post-privatization
pre_data = df[df["category"] == "pre"]
post_data = df[df["category"] == "post"]

# Function to categorize sentiment into Positive, Negative, Neutral
def categorize_sentiment(sentiment):
    if sentiment > 0.1:
        return "Positive"
    elif sentiment < -0.1:
        return "Negative"
    else:
        return "Neutral"

# Group by topic and calculate average sentiment for pre and post
pre_sentiment = pre_data.groupby("Topic Label")
["Sentiment"].mean().reindex(all_topics,
fill_value=0).apply(categorize_sentiment)
post_sentiment = post_data.groupby("Topic Label")
["Sentiment"].mean().reindex(all_topics,
fill_value=0).apply(categorize_sentiment)

# Count articles per topic for pre and post
pre_counts = pre_data["Topic
Label"].value_counts().reindex(all_topics, fill_value=0)
post_counts = post_data["Topic
Label"].value_counts().reindex(all_topics, fill_value=0)

# Combine results into a comparison dataframe
```

```
sentiment_comparison = pd.DataFrame({
    "Topic": all_topics,
    "Pre-Privatization Sentiment": pre_sentiment.values,
    "Post-Privatization Sentiment": post_sentiment.values,
    "Pre Articles Count": pre_counts.values,
    "Post Articles Count": post_counts.values
})

# Display the updated summary
print(sentiment_comparison)

# Save the results
output_file = "pre_post_sentiment_comparison.csv"
sentiment_comparison.to_csv(output_file, index=False)
print(f"Sentiment comparison saved to {output_file}")
```

```
                                    Topic Pre-Privatization Sentiment
\
0      Customer Experience and Operations                     Positive

1   Financial Performance and Investments                      Neutral

2     Domestic Market and Fleet Expansion                     Positive

3         Supply Chain and Sustainability                     Positive

4  Passenger Services and Global Strategy                     Positive


  Post-Privatization Sentiment  Pre Articles Count  Post Articles
Count
0                     Positive                  17
12
1                      Neutral                  36
36
2                     Positive                   8
18
3                     Positive                   8
2
4                     Positive                   1
4
Sentiment comparison saved to pre_post_sentiment_comparison.csv
```

```
# Check sentiment score distribution
print("Sentiment Score Distribution:")
print(df["Sentiment"].describe())

# Plot histogram of sentiment scores
plt.figure(figsize=(10, 6))
sns.histplot(df["Sentiment"], kde=True, bins=20, color="skyblue")
plt.title("Sentiment Score Distribution")
```

```
plt.xlabel("Sentiment Score")
plt.ylabel("Frequency")
plt.axvline(0, color="red", linestyle="--", label="Neutral Threshold")
plt.legend()
plt.show()

Sentiment Score Distribution:
count    142.000000
mean       0.097241
std        0.054115
min       -0.065301
25%        0.069203
50%        0.095331
75%        0.121297
max        0.333838
Name: Sentiment, dtype: float64
```



Graph The sentiment scores are predominantly positive, with most values clustering around 0.1 and a few near the neutral threshold.

```
def categorize_sentiment_adjusted(score):
    if score > 0.1:
        return "Positive"
    elif score < -0.1:
        return "Negative"
```

```
    else:
        return "Neutral"
```

Negative sentiment

```
# Articles with the most negative sentiment
negative_articles = df[df["Sentiment"] < 0].sort_values("Sentiment")
print(negative_articles[["title", "Sentiment", "Processed
Content"]].head(10))
```

```
                                                  title  Sentiment  \
105  Shabby seats, broken IFE. Are upgrades in the ...  -0.065301
36   Centre asks Air India Express to resolve staff...  -0.033840
16   What the fog! Not just bad weather, airlines a...  -0.023267
10   IndiGo to Vistara: Why airlines continue to fa...  -0.013413
8    Air India under the Tatas has a long way to go...  -0.008888

                                      Processed Content
105  runway girl network us cooky using site consen...
36   enjoy additional benefit connect u updated may...
16   listen story january 13 flight mumbai schedule...
10   feedback wee hour 8 may thousand passenger fou...
8    subscribe get feature like last day seen viral...
```

Transformer models for analyzing sentiment

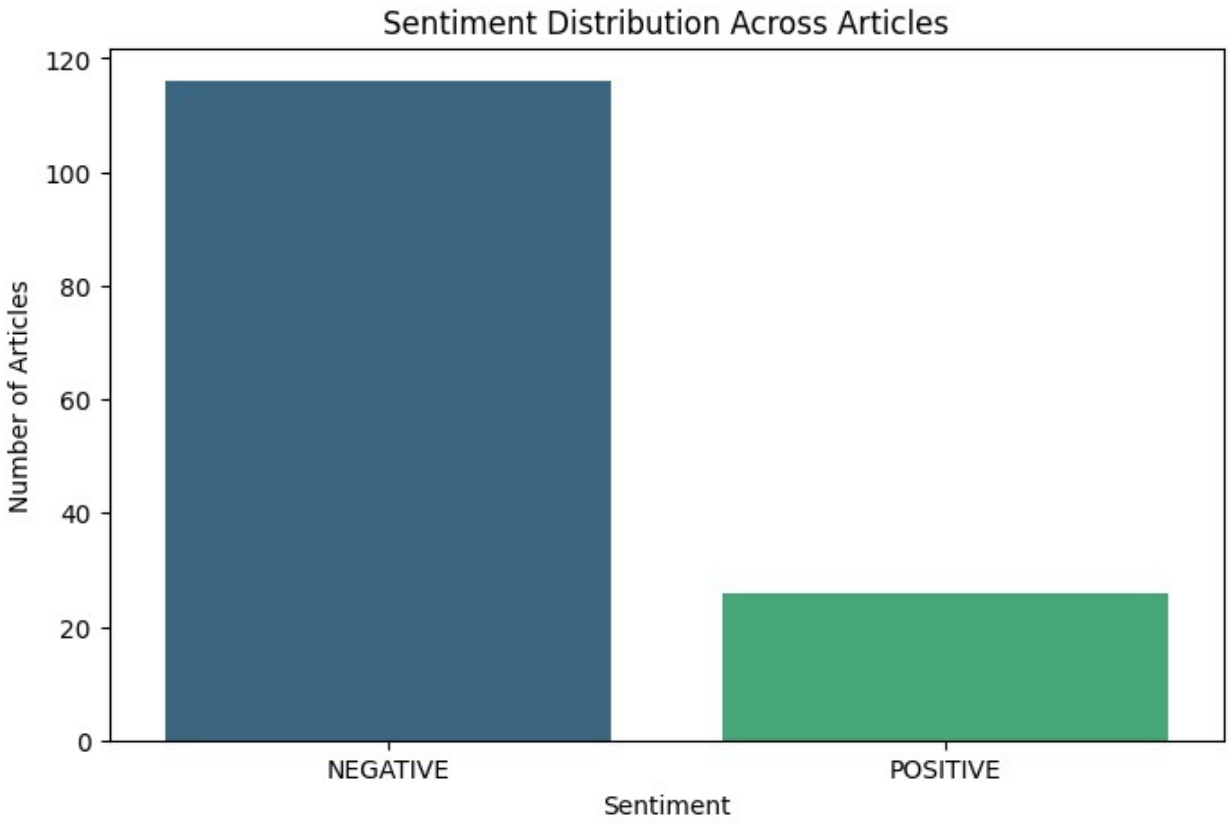
```
from transformers import pipeline
import pandas as pd

# Load the dataset
csv_file = "cleaned_air_india_case_study_articles.csv"
df = pd.read_csv(csv_file)

# Initialize sentiment analysis pipeline
sentiment_analyzer = pipeline("sentiment-analysis", model="distilbert-
base-uncased-finetuned-sst-2-english", truncation=True) #classifies
each article as Positive or Negative

# Truncate the content to 512 tokens
def truncate_text(text, max_length=512):
    return " ".join(text.split()[:max_length])

# Apply truncation and sentiment analysis
print("Performing sentiment analysis with text truncation...")
df["Processed Content"] = df["Processed Content"].apply(lambda x:
truncate_text(x, max_length=512))
df["Transformer Sentiment"] = df["Processed Content"].apply(lambda x:
sentiment_analyzer(x)[0]["label"])
```

```python
# Save the results
output_file = "air_india_with_sentiments.csv"
df.to_csv(output_file, index=False, encoding="utf-8")

print(f"Sentiment analysis complete. Results saved to {output_file}.")
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/
_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your
settings tab (https://huggingface.co/settings/tokens), set it as
secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to
access public models or datasets.
  warnings.warn(

Performing sentiment analysis with text truncation...
Sentiment analysis complete. Results saved to
air_india_with_sentiments.csv.
```

```python
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Load the dataset
df = pd.read_csv("air_india_with_sentiments.csv")

# Plot sentiment distribution
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x="Transformer Sentiment", palette="viridis")
plt.title("Sentiment Distribution Across Articles")
plt.xlabel("Sentiment")
plt.ylabel("Number of Articles")
plt.show()
```

```
<ipython-input-28-24c800b38ffe>:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.countplot(data=df, x="Transformer Sentiment", palette="viridis")
```




```
print(df.head())
```

```
                                                title  \
0  What’s Driving the Privatization of Flag Carri...
1  Why Flag Carriers Are Privatizing in the India...
2  Air India-Vistara Merger: Strategic Win or Tat...
3  Expanding Air India's punctuality woes leave f...
4  India: Is the Tata Group's Air India revamp wo...

                                                 link  \
0  https://airinsight.com/whats-driving-the-priva...
1  https://aviationa2z.com/index.php/2024/11/12/w...
2  https://skift.com/2024/09/17/air-india-vistara...
3  https://timesofindia.indiatimes.com/india/expa...
4  https://www.dw.com/en/india-is-the-tata-groups...

                              date  \
0  11/19/2024, 10:55 PM, +0000 UTC
1  11/12/2024, 05:44 PM, +0000 UTC
2  09/17/2024, 07:00 AM, +0000 UTC
3  09/22/2024, 07:00 AM, +0000 UTC
4  05/16/2024, 07:00 AM, +0000 UTC

                                               query
```

```
aspect  \
0  Air India operational inefficiencies before 2022  Operational
Efficiency
1  Air India operational inefficiencies before 2022  Operational
Efficiency
2  Air India operational inefficiencies before 2022  Operational
Efficiency
3  Air India operational inefficiencies before 2022  Operational
Efficiency
4  Air India operational inefficiencies before 2022  Operational
Efficiency

  category  Publication Year  \
0      pre              2024
1      pre              2024
2      pre              2024
3      pre              2024
4      pre              2024

                                              Content  Word Count  \
0  Commercial Aviation News and Analysis india fl...         758
1  In recent years, the trend of privatizing flag...         691
2  Subscribe today to keep up with the latest tra...         668
3  How to make vrat-friendly Coconut Peanut Chutn...          86
4  Two years after the Tata Group's acquisition o...         900

           Aspect                                    Processed
Content  \
0  Unknown Aspect  commercial aviation news analysis india flag a...

1  Unknown Aspect  recent year trend privatizing flag carrier gai...

2  Unknown Aspect  subscribe today keep latest travel industry ne...

3  Unknown Aspect  make vratfriendly coconut peanut chutney home ...

4  Unknown Aspect  two year tata group acquisition india national...


  Transformer Sentiment
0              NEGATIVE
1              NEGATIVE
2              NEGATIVE
3              POSITIVE
4              NEGATIVE
```

Assigning articles their topics

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
```

```python
# Vectorize the Processed Content for LDA
vectorizer = CountVectorizer(max_df=0.9, min_df=5,
stop_words="english")
content_matrix = vectorizer.fit_transform(df["Processed Content"])

# Fit the LDA model
lda = LatentDirichletAllocation(n_components=5, random_state=42)
lda.fit(content_matrix)

# Assign topics to articles
topic_assignments = lda.transform(content_matrix).argmax(axis=1)

# Assign topics back to the DataFrame
df["Topic"] = topic_assignments

# Define topic labels based on the generated words
topic_labels = {
    0: "Operational Efficiency",
    1: "Customer Experience",
    2: "Financial Performance",
    3: "Fleet Modernization",
    4: "Market Expansion and Competition"
}

# Map topics to human-readable labels
df["Topic Label"] = df["Topic"].map(topic_labels)

print("Topics assigned successfully:")
print(df[["Processed Content", "Topic", "Topic Label"]].head())
```

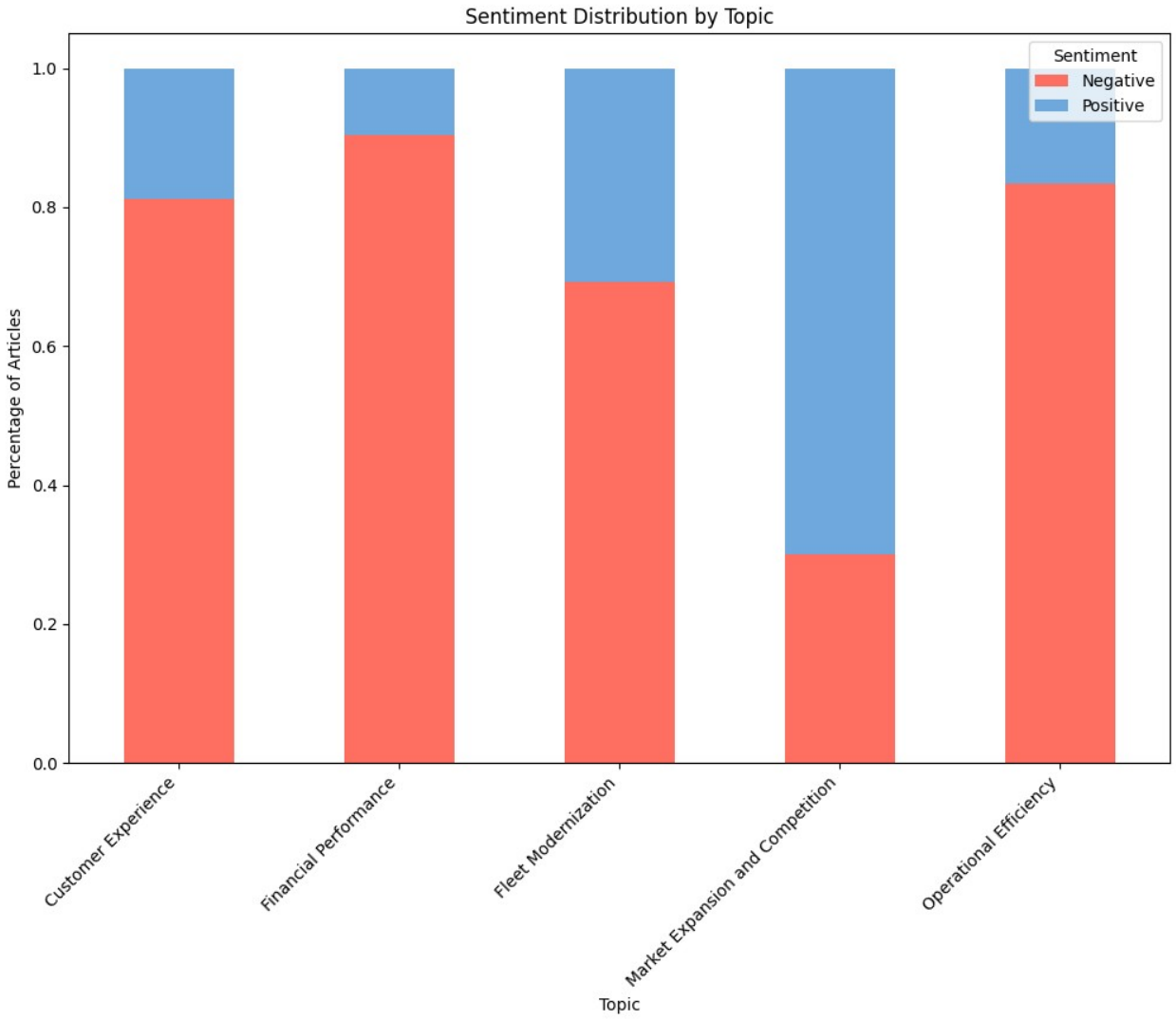
```
Topics assigned successfully:
                                   Processed Content  Topic  \
0  commercial aviation news analysis india flag a...      1
1  recent year trend privatizing flag carrier gai...      1
2  subscribe today keep latest travel industry ne...      2
3  make vratfriendly coconut peanut chutney home ...      3
4  two year tata group acquisition india national...      2

             Topic Label
0    Customer Experience
1    Customer Experience
2  Financial Performance
3    Fleet Modernization
4  Financial Performance
```

```python
# Save the dataset with assigned topics
output_file = "air_india_with_topics.csv"
df.to_csv(output_file, index=False, encoding="utf-8")
print(f"Dataset with topics saved to {output_file}")
```

```
Dataset with topics saved to air_india_with_topics.csv
```

Sentiment Distribution Across Topics

```
# Sentiment Distribution by Topic
topic_sentiment = df.groupby(["Topic Label", "Transformer
Sentiment"]).size().unstack(fill_value=0) # 'unstack' converts the
grouped data into a matrix for easier visualization

# Normalize for percentage distribution
topic_sentiment_percentage =
topic_sentiment.div(topic_sentiment.sum(axis=1), axis=0)

# Plot the distribution
topic_sentiment_percentage.plot(kind='bar', stacked=True, figsize=(12,
8), color=["#FF6F61", "#6FA8DC"])
plt.title("Sentiment Distribution by Topic")
plt.xlabel("Topic")
plt.ylabel("Percentage of Articles")
plt.legend(title="Sentiment", labels=["Negative", "Positive"])
plt.xticks(rotation=45, ha="right")
plt.show()
```




```
print(df.columns)
```

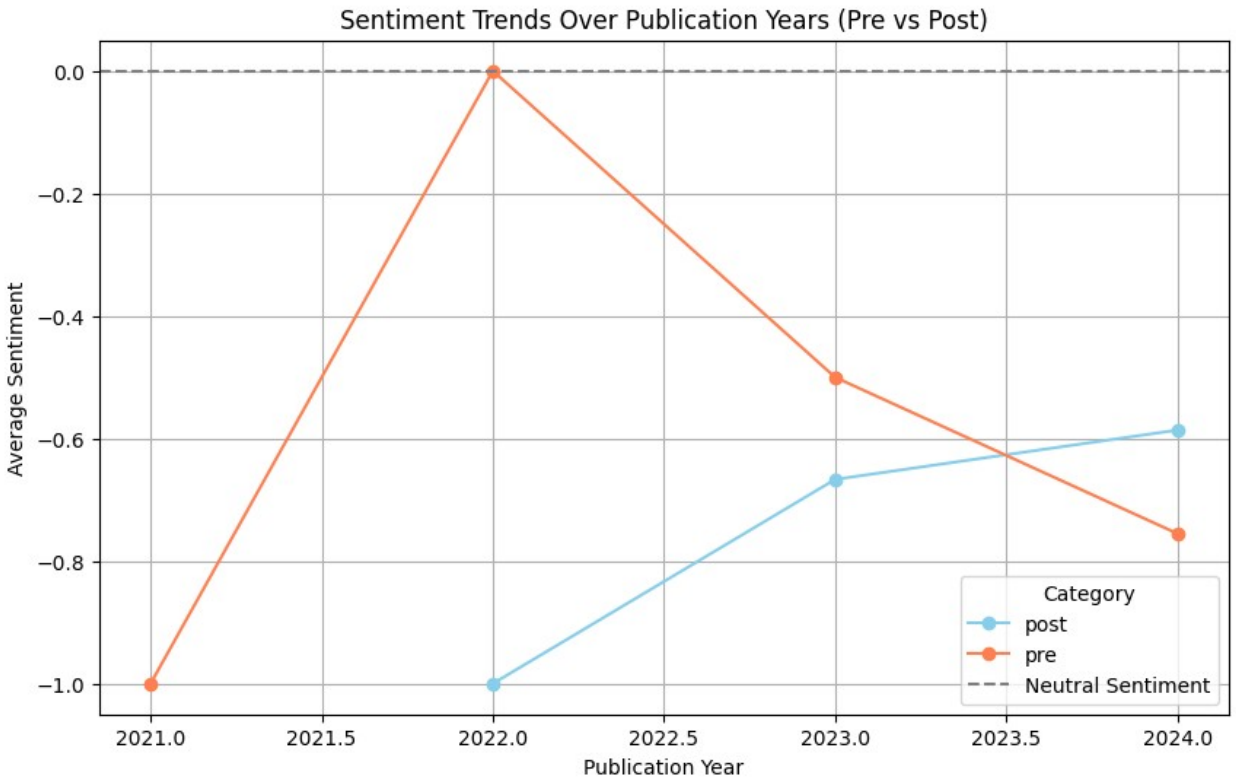
```
Index(['title', 'link', 'date', 'query', 'aspect', 'category',
       'Publication Year', 'Content', 'Word Count', 'Aspect',
       'Processed Content', 'Transformer Sentiment', 'Topic', 'Topic
Label'],
      dtype='object')
```

```
sentiment_mapping = {"POSITIVE": 1, "NEGATIVE": -1}
df["Sentiment"] = df["Transformer Sentiment"].map(sentiment_mapping)
```

Sentiment Trends Over Publication Years

```
sentiment_trends = df.groupby(["Publication Year", "category"])
["Sentiment"].mean().unstack()
sentiment_trends.plot(kind="line", marker="o", figsize=(10, 6),
color=["skyblue", "coral"])
```

```
plt.title("Sentiment Trends Over Publication Years (Pre vs Post)")
plt.xlabel("Publication Year")
plt.ylabel("Average Sentiment")
plt.axhline(0, color="gray", linestyle="--", label="Neutral
Sentiment")
plt.legend(title="Category")
plt.grid(True)
plt.show()
```

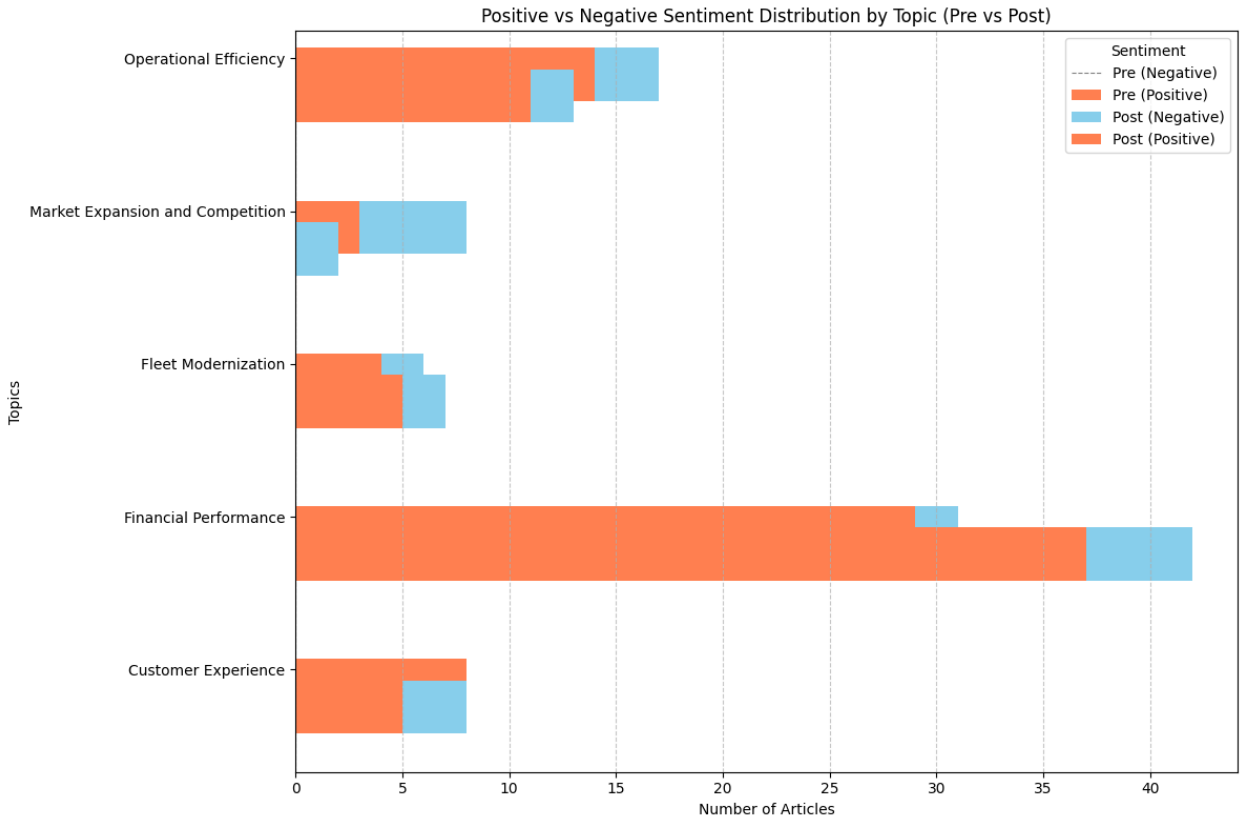


Post-privatization sentiment has steadily improved while pre-privatization sentiment showed high variability and is falling downward.

Distribution of Positive/Negative Sentiment by topic and category.

```
# Prepare data for positive/negative sentiment counts by topic and
category
sentiment_distribution = (
    df.groupby(["Topic Label", "category", "Transformer Sentiment"])
    .size()
    .unstack(fill_value=0)
)

# Plot stacked bar chart
fig, ax = plt.subplots(figsize=(12, 8))
```

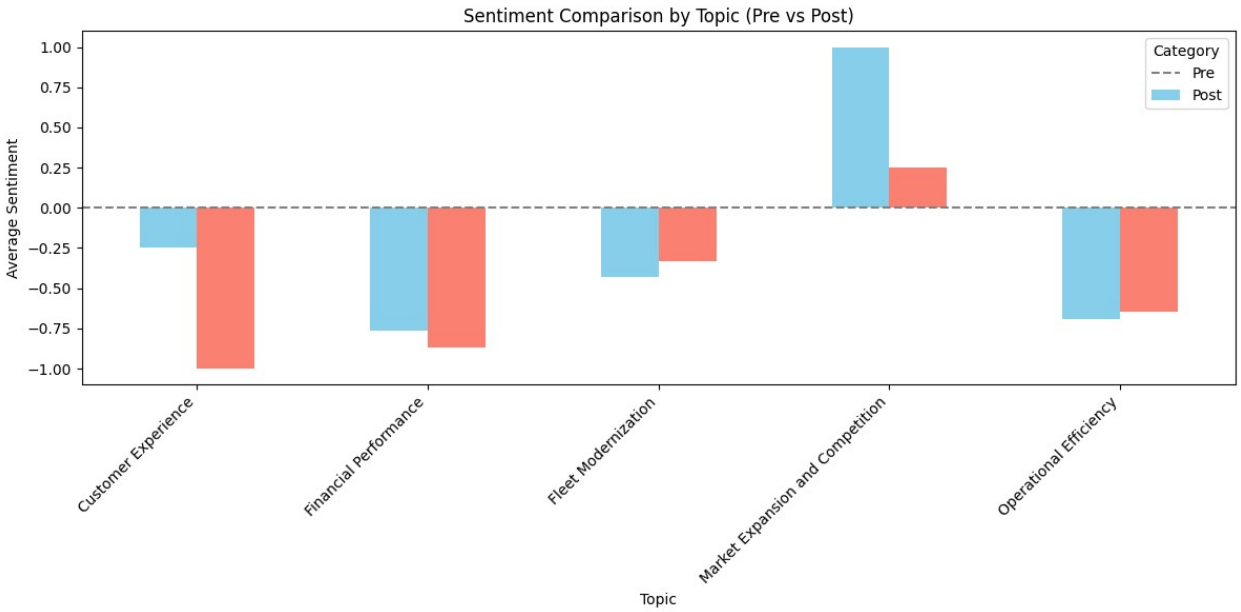
```
# Separate pre- and post-privatization for stacking
pre_data = sentiment_distribution.xs("pre", level="category")
post_data = sentiment_distribution.xs("post", level="category")

# Plot pre-privatization sentiment distribution
pre_data.plot(
    kind="barh",
    stacked=True,
    color={"NEGATIVE": "coral", "POSITIVE": "skyblue"},
    ax=ax,
    label="Pre",
    position=0.8,
    width=0.35,
)

# Plot post-privatization sentiment distribution
post_data.plot(
    kind="barh",
    stacked=True,
    color={"NEGATIVE": "coral", "POSITIVE": "skyblue"},
    ax=ax,
    label="Post",
    position=1.2,
    width=0.35,
)

# Customize plot
ax.set_title("Positive vs Negative Sentiment Distribution by Topic
(Pre vs Post)")
ax.set_xlabel("Number of Articles")
ax.set_ylabel("Topics")
ax.axvline(0, color="gray", linestyle="--", linewidth=0.8)  # Neutral
line
ax.legend(["Pre (Negative)", "Pre (Positive)", "Post (Negative)",
"Post (Positive)"], title="Sentiment")
plt.grid(axis="x", linestyle="--", alpha=0.7)

plt.tight_layout()
plt.show()
```



inancial Performance has more negative articles pre-privatization, while post-privatization shows improved sentiment in several topics like Customer Experience.

Sentiment Comparison by Topic by Category

```
# Group by topic and category to calculate mean sentiment
topic_sentiments = df.groupby(["Topic Label", "category"])
["Sentiment"].mean().unstack()

# Plot the sentiment comparison
topic_sentiments.plot(kind="bar", figsize=(12, 6), color=["skyblue",
"salmon"])
plt.title("Sentiment Comparison by Topic (Pre vs Post)")
plt.xlabel("Topic")
plt.ylabel("Average Sentiment")
plt.axhline(0, color="gray", linestyle="--", label="Neutral
Sentiment")
plt.legend(["Pre", "Post"], title="Category")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()
```




```
# Check a few rows of Processed Content
print(df["Processed Content"].head(10))
```

```
0    commercial aviation news analysis india flag a...
1    recent year trend privatizing flag carrier gai...
2    subscribe today keep latest travel industry ne...
3    make vratfriendly coconut peanut chutney home ...
4    two year tata group acquisition india national...
5    abhishek nayar 02 oct 2024 02 oct 2024 january...
6    marketing performance marketing ecommerce bran...
7    merger air india vistara announced november 20...
8    subscribe get feature like last day seen viral...
9    dubai sideline iata annual general meeting las...
Name: Processed Content, dtype: object
```

```
import spacy

# Load spaCy model
nlp = spacy.load("en_core_web_sm")

# Extract keyphrases using noun chunks
def extract_keyphrases(text):
    doc = nlp(text)
    return [chunk.text for chunk in doc.noun_chunks] # 'noun_chunks'
identifies meaningful contiguous phrases centered around nouns

# Apply the keyphrase extraction
df["Keyphrases"] = df["Processed Content"].apply(extract_keyphrases)

# Preview keyphrases
```

```
print("Keyphrases Extracted with spaCy:")
print(df[["category", "Keyphrases"]].head())
```

```
Keyphrases Extracted with spaCy:
  category                                          Keyphrases
0      pre  [commercial aviation news analysis, india flag...
1      pre  [recent year trend, flag carrier, traction, in...
2      pre  [subscribe, latest travel industry news select...
3      pre  [curry patta plant, faster 10 plant, soil, wat...
4      pre  [two year tata group acquisition india nationa...
```

```
# Group keyphrases by category
pre_keyphrases = df[df["category"] == "pre"]
["Keyphrases"].explode().value_counts() # Explode converts lists in
'Keyphrases' to individual rows for counting
post_keyphrases = df[df["category"] == "post"]
["Keyphrases"].explode().value_counts()

# Combine results into a single DataFrame
keyphrase_comparison = pd.DataFrame({
    "Pre-Privatization": pre_keyphrases,
    "Post-Privatization": post_keyphrases,
}).fillna(0)

print("Keyphrase Comparison (Top 10):")
print(keyphrase_comparison.head(10))

# Save the comparison to CSV
keyphrase_comparison.to_csv("keyphrase_comparison.csv")
print("Keyphrase comparison saved to 'keyphrase_comparison.csv'.")
```

```
Keyphrase Comparison (Top 10):
                                                  Pre-Privatization
\
Keyphrases

01 2024 air india flag carrier india                            0.0

05 percentage point                                             1.0

05 yearonyear increase wizz air                                 1.0

08 2024 1039 ist air transport copyright                        1.0

1 2 mri scan reveal                                             1.0

1 billion colour                                                1.0

1 billion earnings net profit                                   0.0

1 disruptors significant impact demand supply t...              1.0
```

```
1 group                                                          1.0
1 million monthly visitor bhavya                                 6.0

                                                  Post-Privatization
Keyphrases
01 2024 air india flag carrier india                             2.0
05 percentage point                                              0.0
05 yearonyear increase wizz air                                  0.0
08 2024 1039 ist air transport copyright                         0.0
1 2 mri scan reveal                                              0.0
1 billion colour                                                 0.0
1 billion earnings net profit                                    1.0
1 disruptors significant impact demand supply t...               0.0
1 group                                                          0.0
1 million monthly visitor bhavya                                 7.0
Keyphrase comparison saved to 'keyphrase_comparison.csv'.
```

```python
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Generate word clouds
pre_wordcloud = WordCloud(width=800, height=400,
background_color="white").generate(" ".join(pre_keyphrases.index))
post_wordcloud = WordCloud(width=800, height=400,
background_color="white").generate(" ".join(post_keyphrases.index))

# Plot word clouds
plt.figure(figsize=(16, 8))

plt.subplot(1, 2, 1)
plt.imshow(pre_wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Pre-Privatization Keyphrases")

plt.subplot(1, 2, 2)
plt.imshow(post_wordcloud, interpolation="bilinear")
```

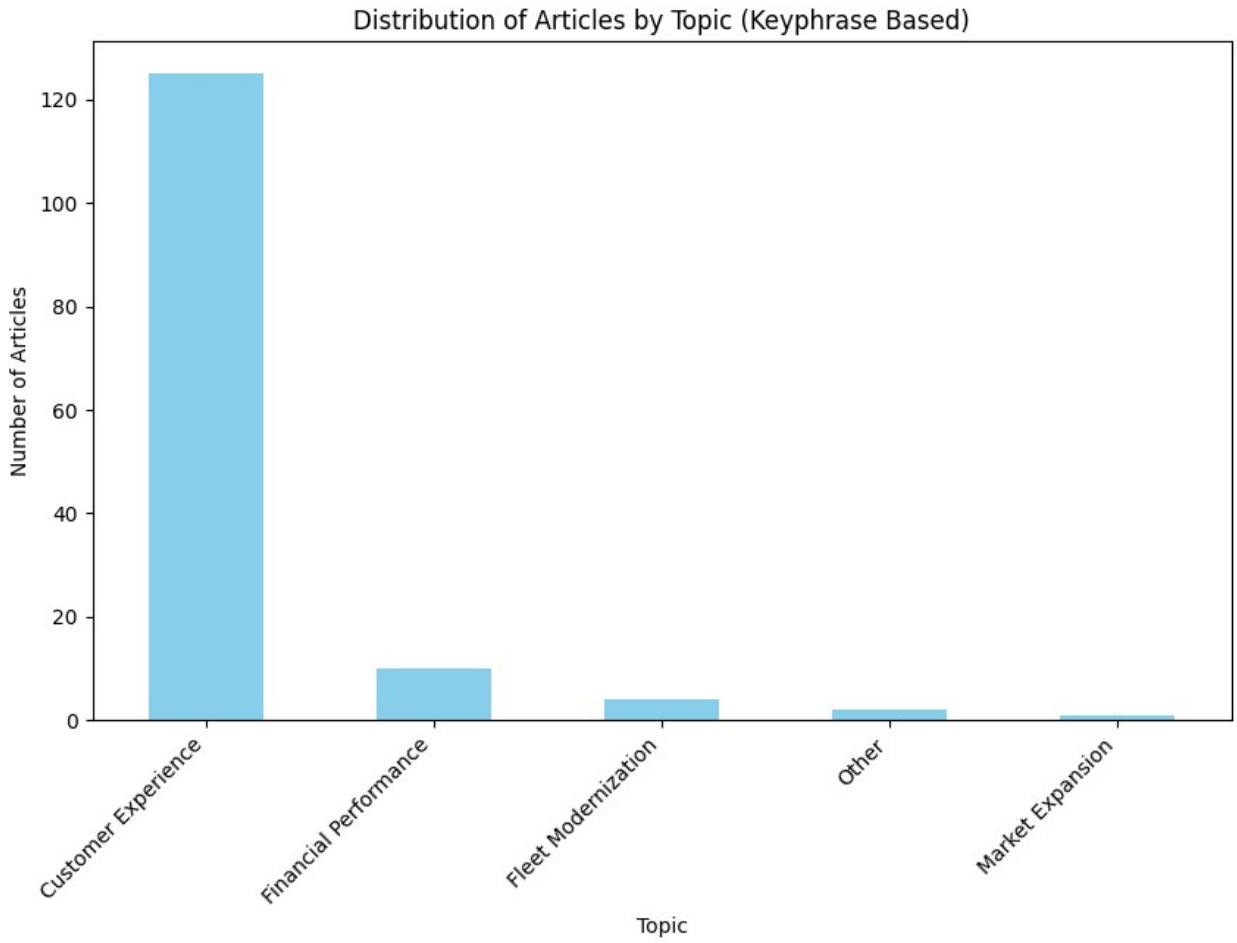
```
plt.axis("off")
plt.title("Post-Privatization Keyphrases")

plt.tight_layout()
plt.show()
```


```
# Example: Assign articles to topics based on keyphrase matches
def assign_topic_from_keyphrases(keyphrases): # Matches keyphrases
against predefined keywords for each topic
    topics = {
        "Customer Experience": ["service", "passenger", "customer",
"experience"],
        "Financial Performance": ["financial", "profit",
"investment"],
        "Fleet Modernization": ["aircraft", "fleet", "modernization",
"order"],
        "Market Expansion": ["market", "route", "global"],
        "Operational Efficiency": ["operational", "efficiency",
"cost"]
    }
    for topic, keywords in topics.items():
        if any(keyword in " ".join(keyphrases) for keyword in
keywords):
            return topic
    return "Other"

df["Topic (Keyphrase Based)"] =
df["Keyphrases"].apply(assign_topic_from_keyphrases)

print("Updated Dataset with Topics:")
print(df[["category", "Topic (Keyphrase Based)"]].head())
```

```
Updated Dataset with Topics:
  category Topic (Keyphrase Based)
0      pre   Financial Performance
1      pre     Customer Experience
2      pre     Customer Experience
3      pre                   Other
4      pre     Customer Experience
```

Distribution of Articles by Topic

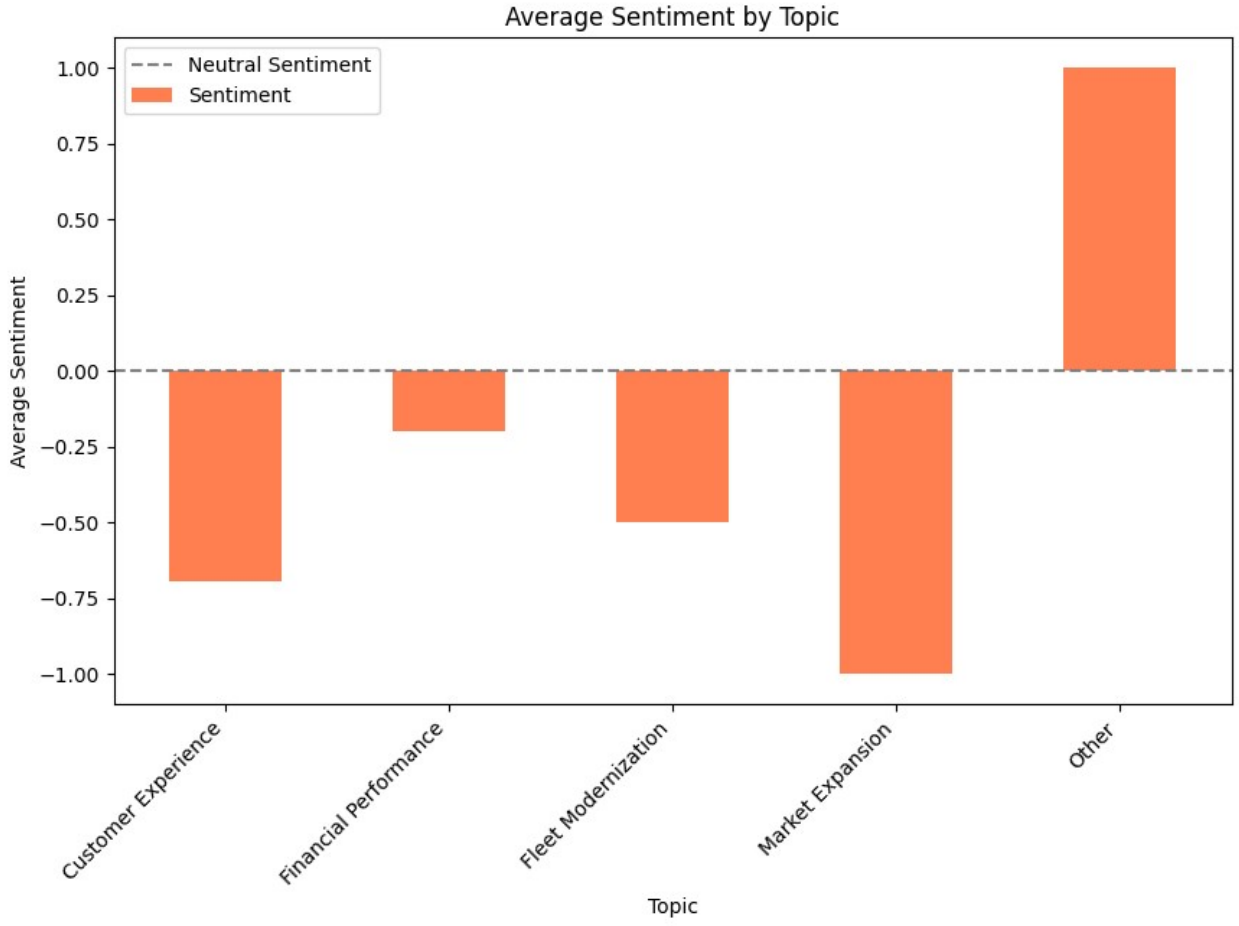
```
# Count articles by topic
topic_distribution = df["Topic (Keyphrase Based)"].value_counts()

# Plot topic distribution
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
topic_distribution.plot(kind="bar", color="skyblue")
plt.title("Distribution of Articles by Topic (Keyphrase Based)")
plt.xlabel("Topic")
plt.ylabel("Number of Articles")
plt.xticks(rotation=45, ha="right")
plt.show()
```



Most articles are related to customer experience, indicating it is a primary focus. "Market Expansion" and "Fleet Modernization" have significantly fewer articles, showing less focus or coverage.

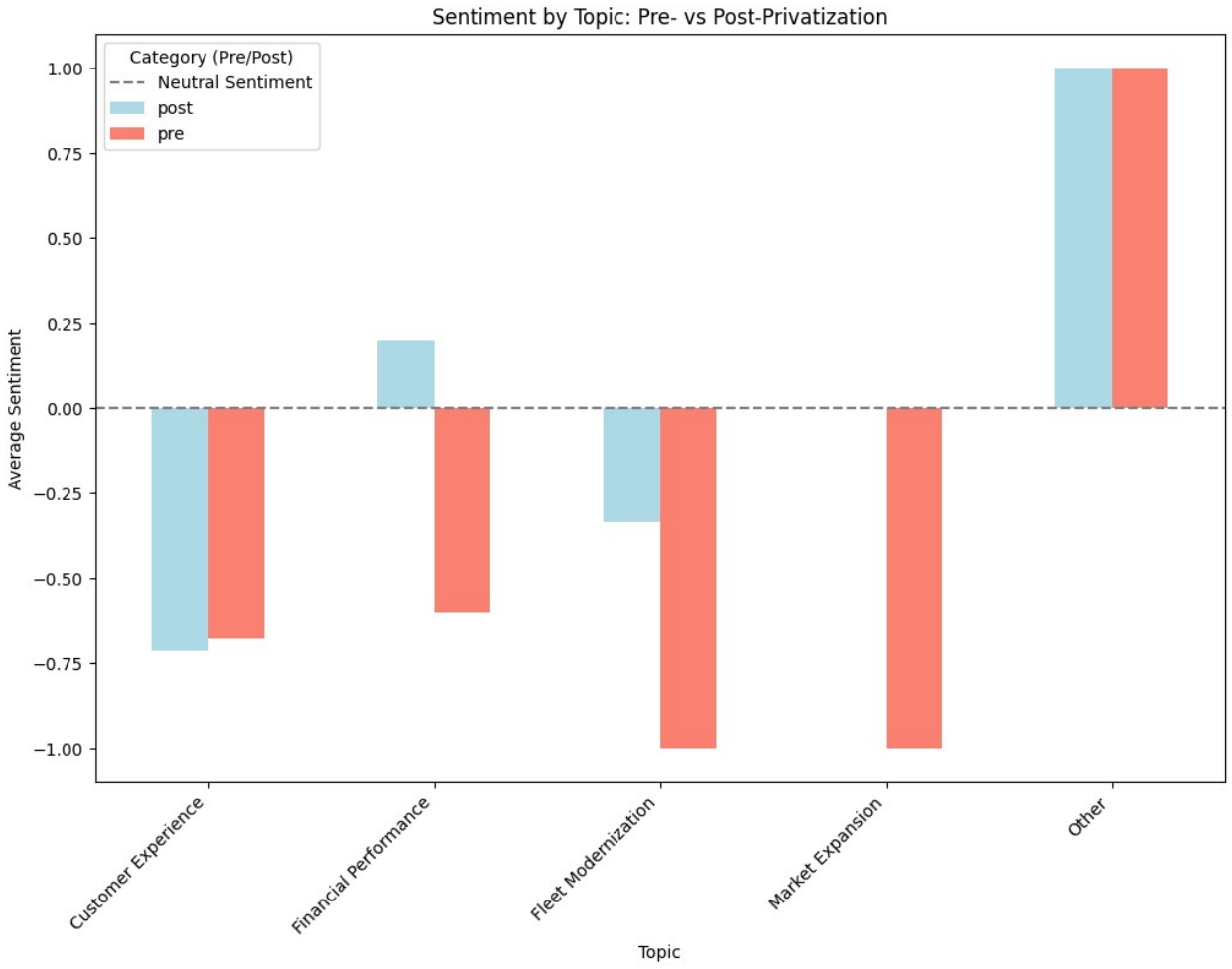
Average Sentiment by Topic

```
# Group by topic and calculate average sentiment
topic_sentiment = df.groupby("Topic (Keyphrase Based)")
["Sentiment"].mean()

# Plot topic sentiment
plt.figure(figsize=(10, 6))
topic_sentiment.plot(kind="bar", color="coral")
plt.title("Average Sentiment by Topic")
plt.xlabel("Topic")
plt.ylabel("Average Sentiment")
plt.axhline(0, color="gray", linestyle="--", label="Neutral
Sentiment")
plt.legend()
plt.xticks(rotation=45, ha="right")
plt.show()
```



Topics like "Customer Experience" and "Market Expansion" have negative average sentiment scores, indicating dissatisfaction or criticism.Topics like "Other" show a positive sentiment, potentially reflecting optimism or positive feedback.

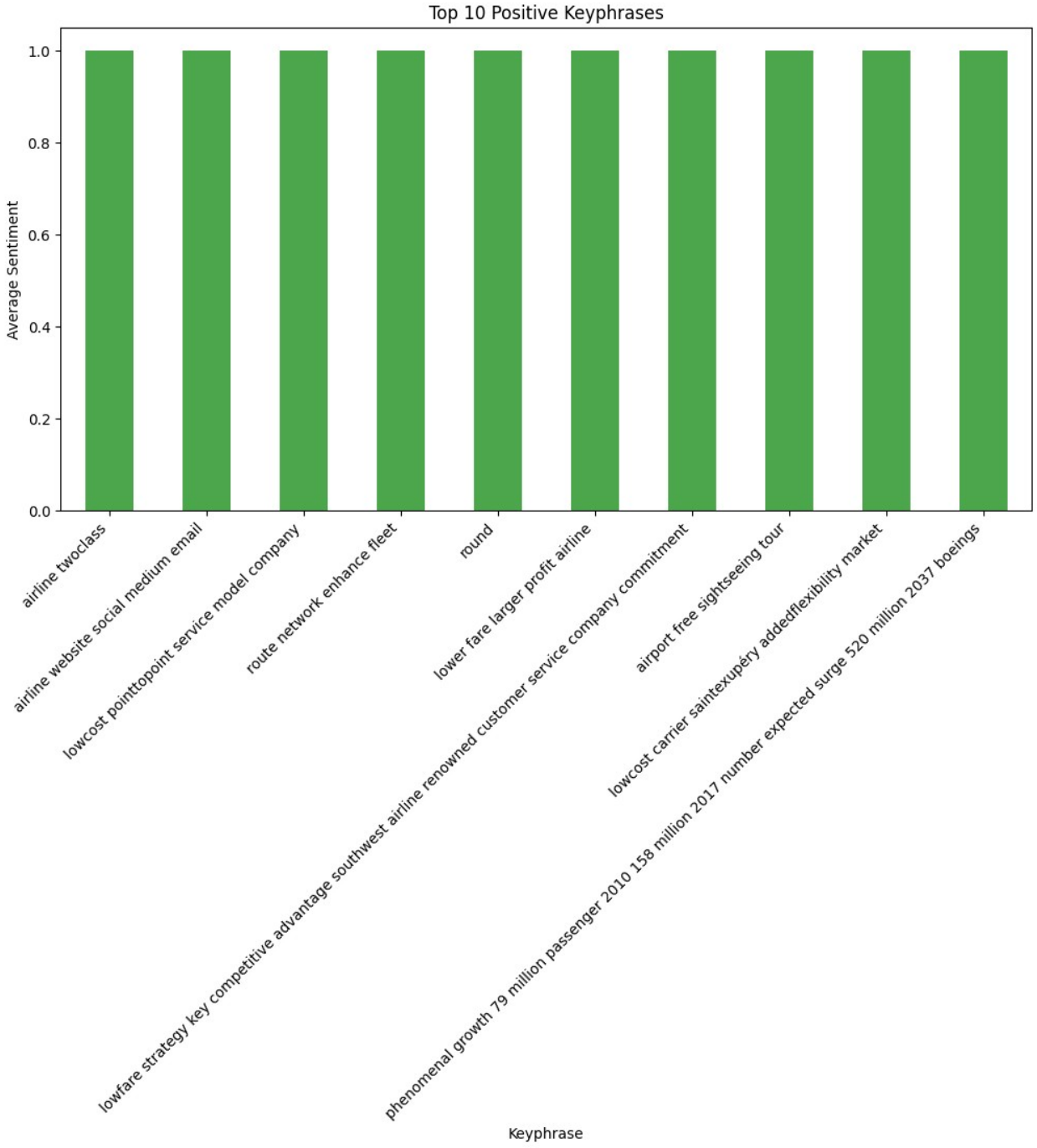
```
# Group by category and topic to calculate average sentiment
pre_post_sentiment = df.groupby(["category", "Topic (Keyphrase
Based)"])["Sentiment"].mean().unstack()

# Plot pre vs post sentiment by topic
pre_post_sentiment.T.plot(kind="bar", figsize=(12, 8),
color=["lightblue", "salmon"])
plt.title("Sentiment by Topic: Pre- vs Post-Privatization")
plt.xlabel("Topic")
plt.ylabel("Average Sentiment")
plt.axhline(0, color="gray", linestyle="--", label="Neutral
Sentiment")
plt.legend(title="Category (Pre/Post)")
plt.xticks(rotation=45, ha="right")
plt.show()
```

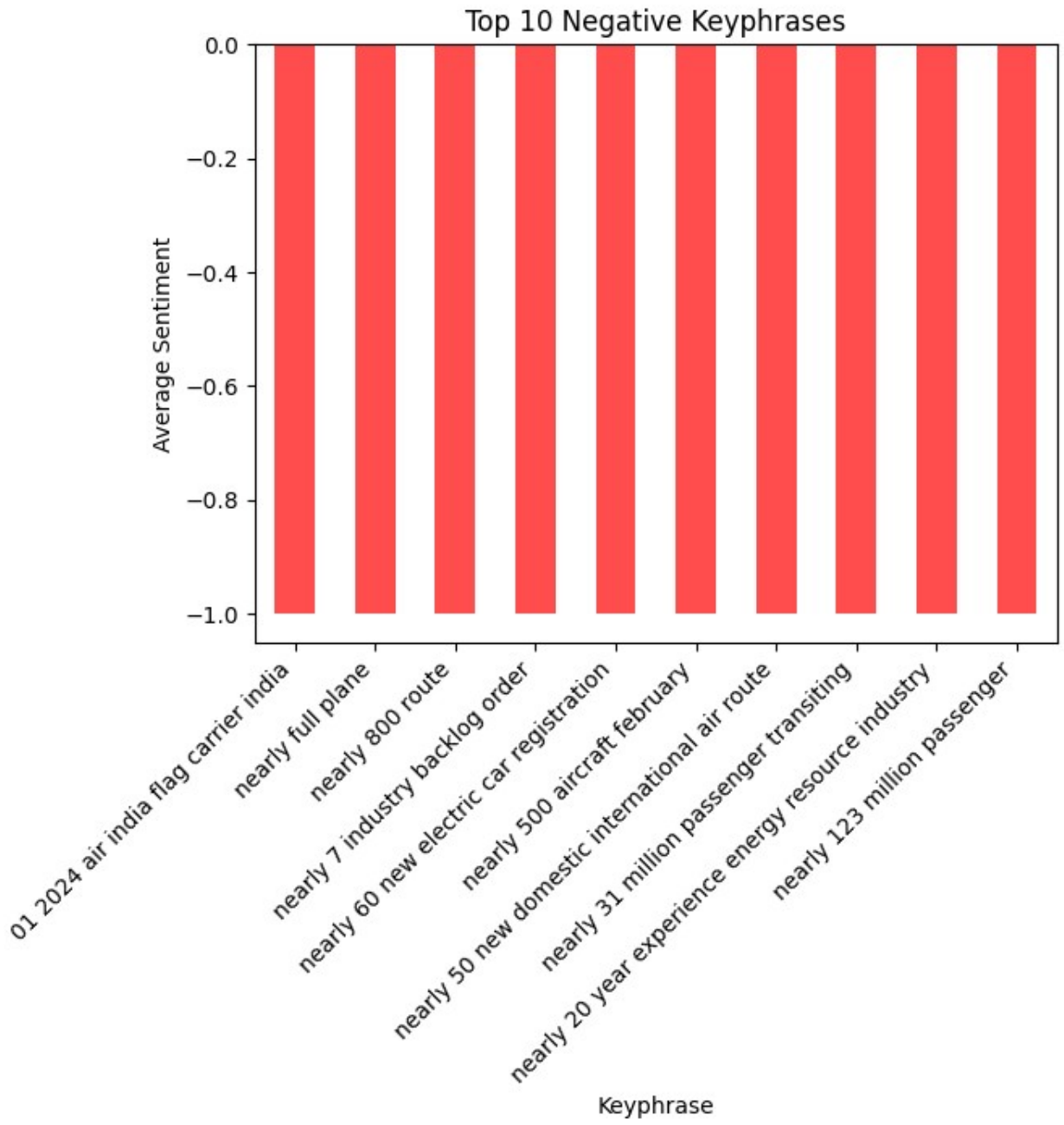


Negative sentiment is dominant for topics like "Customer Experience" and "Market Expansion." Improved sentiment for topics like "Financial Performance" and "Fleet Modernization."

Top 10 Positive/Negative Keyphrases

```
# Sentiment distribution by keyphrase
keyphrase_sentiments = df.explode("Keyphrases").groupby("Keyphrases")
["Sentiment"].mean().sort_values()

# Plot top positive and negative keyphrases
plt.figure(figsize=(12, 6))

# Top 10 positive keyphrases
keyphrase_sentiments.tail(10).plot(kind="bar", color="green",
alpha=0.7)
plt.title("Top 10 Positive Keyphrases")
plt.xlabel("Keyphrase")
plt.ylabel("Average Sentiment")
plt.xticks(rotation=45, ha="right")
plt.show()

# Top 10 negative keyphrases
keyphrase_sentiments.head(10).plot(kind="bar", color="red", alpha=0.7)
plt.title("Top 10 Negative Keyphrases")
plt.xlabel("Keyphrase")
plt.ylabel("Average Sentiment")
plt.xticks(rotation=45, ha="right")
plt.show()
```

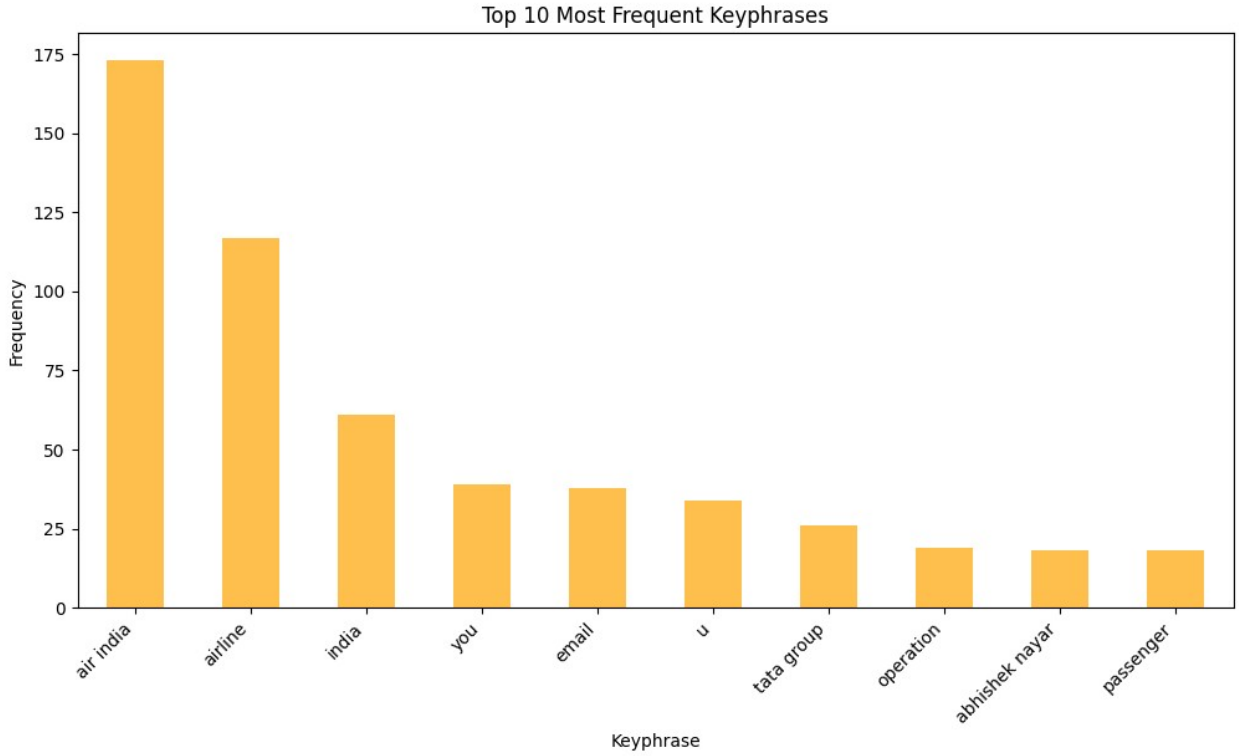

Top 10 Positive Keyphrases
Average Sentiment
1.0
0.8
0.6
0.4
0.2
0.0
airline twoclass
airline website social medium email
lowcost pointtopoint service model company
route network enhance fleet
round
lower fare larger profit airline
lowfare strategy key competitive advantage southwest airline renowned customer service company commitment
airport free sightseeing tour
lowcost carrier saintexupéry addedflexibility market
phenomenal growth 79 million passenger 2010 158 million 2017 number expected surge 520 million 2037 boeings
Keyphrase



The green bar chart indicates keyphrases that are associated with the most positive sentiments in the dataset. The red bar chart highlights keyphrases that are associated with the most negative sentiments.

```
# Frequency of keyphrases
keyphrase_frequency = df.explode("Keyphrases")
["Keyphrases"].value_counts()

# Top 10 keyphrases
plt.figure(figsize=(12, 6))
```

```
keyphrase_frequency.head(10).plot(kind="bar", color="orange",
alpha=0.7)
plt.title("Top 10 Most Frequent Keyphrases")
plt.xlabel("Keyphrase")
plt.ylabel("Frequency")
plt.xticks(rotation=45, ha="right")
plt.show()
```



Keyphrases like "air india" and "airline" appear the most frequently, highlighting their centrality to the dataset. The chart highlights the relative importance or focus of different terms in the dataset.

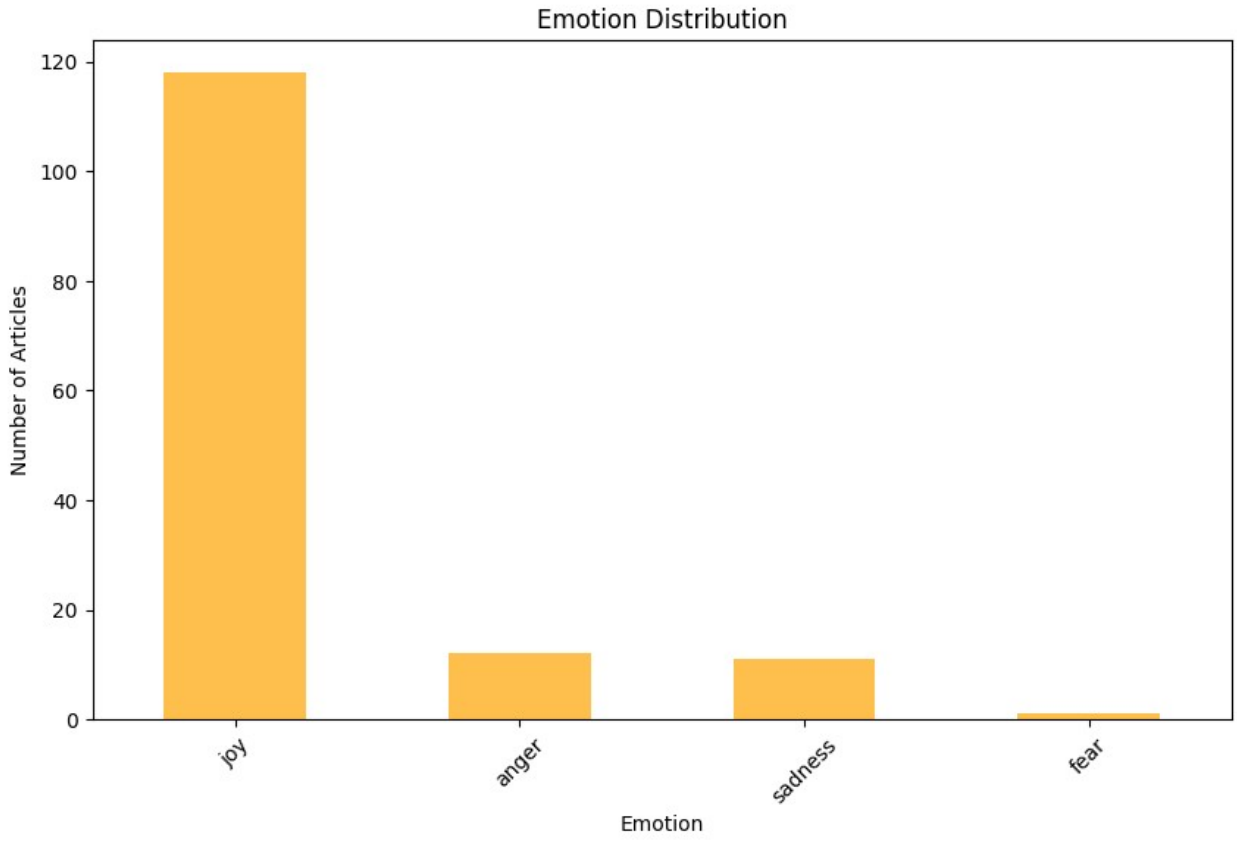
```
from wordcloud import WordCloud

# Positive and negative keyphrases
positive_keyphrases = " ".join(df[df["Sentiment"] > 0]
["Keyphrases"].explode().dropna())
negative_keyphrases = " ".join(df[df["Sentiment"] < 0]
["Keyphrases"].explode().dropna())

# Generate word clouds
plt.figure(figsize=(12, 6))
wc_positive = WordCloud(background_color="white",
colormap="Greens").generate(positive_keyphrases)
wc_negative = WordCloud(background_color="white",
colormap="Reds").generate(negative_keyphrases)
```

```
plt.subplot(1, 2, 1)
plt.imshow(wc_positive, interpolation="bilinear")
plt.axis("off")
plt.title("Positive Keyphrases")

plt.subplot(1, 2, 2)
plt.imshow(wc_negative, interpolation="bilinear")
plt.axis("off")
plt.title("Negative Keyphrases")

plt.show()
```

Green word cloud showing the most frequently occurring keyphrases in positively scored articles. Red word cloud displaying the keyphrases in negatively scored articles.

```
from transformers import AutoTokenizer, pipeline
import matplotlib.pyplot as plt

# Load the emotion classifier and its tokenizer
tokenizer = AutoTokenizer.from_pretrained("bhadresh-savani/distilbert-
base-uncased-emotion")
emotion_classifier = pipeline("text-classification", model="bhadresh-
savani/distilbert-base-uncased-emotion", tokenizer=tokenizer)

# Define a function to truncate text using the tokenizer
def truncate_text_for_model(text, max_length=512):
    encoded = tokenizer.encode(text, truncation=True,
max_length=max_length, return_tensors="pt")
    return tokenizer.decode(encoded[0], skip_special_tokens=True)

# Truncate "Processed Content" to fit within 512 tokens
df["Truncated Content"] = df["Processed Content"].apply(lambda x:
truncate_text_for_model(x, max_length=512))

# Predict emotions on truncated content
df["Emotion"] = df["Truncated Content"].apply(
    lambda x: emotion_classifier(x)[0]["label"]
)
```

```
# Visualize emotion distribution
plt.figure(figsize=(10, 6))
df["Emotion"].value_counts().plot(kind="bar", color="orange",
alpha=0.7)
plt.title("Emotion Distribution")
plt.xlabel("Emotion")
plt.ylabel("Number of Articles")
plt.xticks(rotation=45)
plt.show()
```



Joy is the dominant emotion, appearing in the majority of articles. Anger, sadness are present but occur less frequently. Fear Indicates minimal association of fear-related content in the dataset

LLM

```
pip install openai

Requirement already satisfied: openai in
/usr/local/lib/python3.10/dist-packages (1.54.4)
Requirement already satisfied: anyio<5,>=3.5.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in
```

```
/usr/local/lib/python3.10/dist-packages (from openai) (1.9.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (0.27.2)
Requirement already satisfied: jiter<1,>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (0.7.1)
Requirement already satisfied: pydantic<3,>=1.9.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (2.9.2)
Requirement already satisfied: sniffio in
/usr/local/lib/python3.10/dist-packages (from openai) (1.3.1)
Requirement already satisfied: tqdm>4 in
/usr/local/lib/python3.10/dist-packages (from openai) (4.66.6)
Requirement already satisfied: typing-extensions<5,>=4.11 in
/usr/local/lib/python3.10/dist-packages (from openai) (4.12.2)
Requirement already satisfied: idna>=2.8 in
/usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai)
(3.10)
Requirement already satisfied: exceptiongroup in
/usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai)
(1.2.2)
Requirement already satisfied: certifi in
/usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0-
>openai) (2024.8.30)
Requirement already satisfied: httpcore==1.* in
/usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0-
>openai) (1.0.7)
Requirement already satisfied: h11<0.15,>=0.13 in
/usr/local/lib/python3.10/dist-packages (from httpcore==1.*-
>httpx<1,>=0.23.0->openai) (0.14.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0-
>openai) (0.7.0)
Requirement already satisfied: pydantic-core==2.23.4 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0-
>openai) (2.23.4)

pip install openai pandas

Requirement already satisfied: openai in
/usr/local/lib/python3.10/dist-packages (1.54.4)
Requirement already satisfied: pandas in
/usr/local/lib/python3.10/dist-packages (2.2.2)
Requirement already satisfied: anyio<5,>=3.5.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (1.9.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (0.27.2)
Requirement already satisfied: jiter<1,>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from openai) (0.7.1)
Requirement already satisfied: pydantic<3,>=1.9.0 in
```

```
/usr/local/lib/python3.10/dist-packages (from openai) (2.9.2)
Requirement already satisfied: sniffio in
/usr/local/lib/python3.10/dist-packages (from openai) (1.3.1)
Requirement already satisfied: tqdm>4 in
/usr/local/lib/python3.10/dist-packages (from openai) (4.66.6)
Requirement already satisfied: typing-extensions<5,>=4.11 in
/usr/local/lib/python3.10/dist-packages (from openai) (4.12.2)
Requirement already satisfied: numpy>=1.22.4 in
/usr/local/lib/python3.10/dist-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
Requirement already satisfied: idna>=2.8 in
/usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai)
(3.10)
Requirement already satisfied: exceptiongroup in
/usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai)
(1.2.2)
Requirement already satisfied: certifi in
/usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0-
>openai) (2024.8.30)
Requirement already satisfied: httpcore==1.* in
/usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0-
>openai) (1.0.7)
Requirement already satisfied: h11<0.15,>=0.13 in
/usr/local/lib/python3.10/dist-packages (from httpcore==1.*-
>httpx<1,>=0.23.0->openai) (0.14.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0-
>openai) (0.7.0)
Requirement already satisfied: pydantic-core==2.23.4 in
/usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0-
>openai) (2.23.4)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas) (1.16.0)

import openai
import pandas as pd

# Set your OpenAI API key
openai.api_key = "sk-proj-MmfIwlaVrJu7Ulb1DXPYgS-
4Iya78f23CXvZwxrKJNfNadTOyxWaqXiQLD0cRP-EKK2n8GzLs-
T3BlbkFJFedsAxKFxY59P2sNJHjvYdfvqCJELg5q16UHrojeftbu-
LxlvvpBlBQ2AnBiDwHm3gd_322_EA"

# Define a cost-efficient model and token limit
```

```
MODEL = "gpt-3.5-turbo"
MAX_TOKENS = 150  # Limit tokens per query to control costs

# Load your cleaned dataset
df = pd.read_csv('cleaned_air_india_case_study_articles.csv')

# Display the first few rows for verification
print(df.head())
```

```
                                                title  \
0  What’s Driving the Privatization of Flag Carri...
1  Why Flag Carriers Are Privatizing in the India...
2  Air India-Vistara Merger: Strategic Win or Tat...
3  Expanding Air India's punctuality woes leave f...
4  India: Is the Tata Group's Air India revamp wo...

                                                 link  \
0  https://airinsight.com/whats-driving-the-priva...
1  https://aviationa2z.com/index.php/2024/11/12/w...
2  https://skift.com/2024/09/17/air-india-vistara...
3  https://timesofindia.indiatimes.com/india/expa...
4  https://www.dw.com/en/india-is-the-tata-groups...

                               date  \
0  11/19/2024, 10:55 PM, +0000 UTC
1  11/12/2024, 05:44 PM, +0000 UTC
2  09/17/2024, 07:00 AM, +0000 UTC
3  09/22/2024, 07:00 AM, +0000 UTC
4  05/16/2024, 07:00 AM, +0000 UTC

                                              query
aspect  \
0  Air India operational inefficiencies before 2022  Operational
Efficiency
1  Air India operational inefficiencies before 2022  Operational
Efficiency
2  Air India operational inefficiencies before 2022  Operational
Efficiency
3  Air India operational inefficiencies before 2022  Operational
Efficiency
4  Air India operational inefficiencies before 2022  Operational
Efficiency

  category  Publication Year  \
0      pre              2024
1      pre              2024
2      pre              2024
3      pre              2024
4      pre              2024
```

```
                                             Content  Word Count  \
0  Commercial Aviation News and Analysis india fl...         758
1  In recent years, the trend of privatizing flag...         691
2  Subscribe today to keep up with the latest tra...         668
3  How to make vrat-friendly Coconut Peanut Chutn...          86
4  Two years after the Tata Group's acquisition o...         900

            Aspect                                  Processed Content
0  Unknown Aspect  commercial aviation news analysis india flag a...
1  Unknown Aspect  recent year trend privatizing flag carrier gai...
2  Unknown Aspect  subscribe today keep latest travel industry ne...
3  Unknown Aspect  make vratfriendly coconut peanut chutney home ...
4  Unknown Aspect  two year tata group acquisition india national...
```

```
pip install openai==0.27.8
```

```
Collecting openai==0.27.8
  Downloading openai-0.27.8-py3-none-any.whl.metadata (13 kB)
Requirement already satisfied: requests>=2.20 in
/usr/local/lib/python3.10/dist-packages (from openai==0.27.8) (2.32.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-
packages (from openai==0.27.8) (4.66.6)
Requirement already satisfied: aiohttp in
/usr/local/lib/python3.10/dist-packages (from openai==0.27.8) (3.11.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.20-
>openai==0.27.8) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.20-
>openai==0.27.8) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.20-
>openai==0.27.8) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.20-
>openai==0.27.8) (2024.8.30)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(2.4.3)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(1.3.1)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(24.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
```

```
(6.1.0)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(0.2.0)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(1.17.2)
Requirement already satisfied: async-timeout<6.0,>=4.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->openai==0.27.8)
(4.0.3)
Requirement already satisfied: typing-extensions>=4.1.0 in
/usr/local/lib/python3.10/dist-packages (from multidict<7.0,>=4.5-
>aiohttp->openai==0.27.8) (4.12.2)
Downloading openai-0.27.8-py3-none-any.whl (73 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0.0/73.6 kB ? eta -:--:--
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 73.6/73.6 kB 5.6 MB/s eta
0:00:00
pting uninstall: openai
    Found existing installation: openai 1.54.4
    Uninstalling openai-1.54.4:
      Successfully uninstalled openai-1.54.4
Successfully installed openai-0.27.8
```


```
{"id":"41849876fcb04604ab456e3726938e40","pip_warning":{"packages":
["openai"]}}
```


```python
import openai
import pandas as pd

# Set your OpenAI API key
openai.api_key = "sk-proj-MmfIwlaVrJu7Ulb1DXPYgS-
4Iya78f23CXvZwxrKJNfNadTOyxWaqXiQLD0cRP-EKK2n8GzLs-
T3BlbkFJFedsAxKFxY59P2sNJHjvYdfvqCJELg5q16UHrojeftbu-
LxlvvpBlBQ2AnBiDwHm3gd_322_EA"


# Define the Summarization Function
def summarize_article(article_text):
    try:
        response = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system", "content": "You are an expert
summarizer."},
                {"role": "user", "content": f"Summarize the following
article in 100 words or less: {article_text}"},
            ],
            max_tokens=150,  # Limit response length
            temperature=0.5  # Lower temperature for deterministic
results
```

```
        )
        # Access the generated summary from the response object
        summary = response['choices'][0]['message']['content'].strip()
        return summary
    except Exception as e:
        return f"Error: {str(e)}"

# Load the dataset (adjust the file path and column name)
df = pd.read_csv('cleaned_air_india_case_study_articles.csv')

# Ensure the 'Content' column exists and has valid data
if 'Content' not in df.columns or df['Content'].isnull().all():
    raise ValueError("The 'Content' column is missing or empty in the
dataset.")

# Test the function on a single article
sample_article = df['Content'][0]  # Replace with your column name if
different
summary = summarize_article(sample_article)

print("Sample Article:")
print(sample_article)
print("\nGenerated Summary:")
print(summary)
```

```
Sample Article:
Commercial Aviation News and Analysis india flag a4 Over the past
decade, the Indian subcontinent has witnessed a growing shift toward
privatizing national flag carriers. Air India, Pakistan International
Airlines (PIA), and SriLankan Airlines have been at the center of
these efforts, driven by a need to address mounting financial losses
and operational inefficiencies. Historically, these airlines
represented national pride, serving as emblems of their countries on
the global stage. However, the push for privatization stems from
various systemic issues, including excessive political interference,
persistent debt, and inadequate management. Let’s explore each
airline’s unique challenges and why privatization may offer a path
forward. Once synonymous with luxury in Indian aviation, Air India
gradually became a symbol of inefficiency and debt over the decades.
By 2020, the Indian flag carrier’s financial troubles had reached
critical levels, with debts hovering around $8 billion, as reported by
Aviationa2z. Despite years of attempted bailouts, the airline’s
operational inefficiencies, including overstaffing and routes failing
to profit, only worsened its financial situation. The persistent
losses and high-interest debt ultimately led the Indian government to
look for private ownership as a lifeline. In 2022, the Tata Group
acquired Air India. This acquisition marked a historic turn in Indian
aviation, as Tata initiated a series of reforms to rejuvenate the
carrier. Since the acquisition, the airline has placed substantial
aircraft orders and set ambitious plans to revamp its fleet. Moreover,
```

Air India also merged Vistara and Air India Express into its operations, signaling its intent to make Air India a major global airline. Pakistan International Airlines (PIA) has also been suffering from issues ranging from political interference to massive debt. Allegations of corruption, coupled with overstaffing and mismanagement, have further affected the airline's financial woes. Political appointments and frequent leadership changes have further hampered PIA's ability to establish a consistent, long-term strategy, resulting in estimated losses nearing $5 billion. Facing these challenges, the Pakistani government announced plans to privatize PIA, hoping to curb the recurring financial losses. However, attracting investors has proven difficult, with potential buyers wary of the airline's persistent inefficiencies and high debt burden. For PIA, privatization represents a daunting yet necessary step to restore financial health, even if significant restructuring may be required beforehand. SriLankan Airlines has faced similar challenges, but its journey toward privatization hit a major blockade in 2016. The airline, struggling due to unsustainable debt and management inefficiencies, was on the privatization path as the Sri Lankan government sought to relieve the financial pressure. However, political instability and limited investor interest distracted the effort. The airline's unattractive financial profile and inconsistent political commitment have since stalled further privatization plans. The story of SriLankan Airlines serves as a cautionary tale on how investor hesitation and political instability can affect even the best-intentioned privatization efforts. Despite ongoing discussions, finding willing investors has remained a challenge. The path to privatization in South Asia comes with numerous shared challenges. Here's a closer look at the common issues: While privatization isn't a one-size-fits-all solution, it offers a chance for these airlines to realign priorities and operate independently of political pressures. Air India's transition under the Tata Group is an early example of the potential benefits of professional management, improved accountability, and enhanced efficiency. Still, privatization has risks and requires government support to succeed. Without a commitment to minimize political interference and allow operational freedom, the benefits of private ownership may be limited. For PIA and Sri Lankan Airlines, achieving such autonomy remains an uphill challenge. The drive to privatize flag carriers in South Asia highlights the limitations of state ownership due to financial and political burdens. While Air India's privatization has shown early promise, the paths for PIA and Sri Lankan Airlines are less clear-cut and are shaped by distinct national challenges. What remains evident is that for these carriers to thrive, long-term commitments to restructuring and operational freedom are crucial. What are your thoughts on this trend and the issues behind it? Let us know in the comments below. Views: 31 Your email address will not be published. Required fields are marked * Comment * Name Email Website Δ This site uses Akismet to reduce spam. Learn how your comment data is processed. 717 737 MAX 787 2024 A320NEO

```
Aerospace AIR Airbus Airbus A321neo airlines airlines’ American
Airlines and aviation Boeing Bombardier Business Aviation COMAC
Commercial Aviation CSeries daily deliveries Delta Air Lines Embraer
FAA for GTF India Insight JetBlue key Latest News Lufthansa Max
Mitsubishi morning call orders Pratt & Whitney Rolls-Royce social
media Southwest Airlines stories the United Airlines Urban Air
Mobility

Generated Summary:
The article discusses the trend of privatizing national flag carriers
in the Indian subcontinent, focusing on Air India, Pakistan
International Airlines (PIA), and SriLankan Airlines. These airlines
have faced financial losses and operational inefficiencies, leading to
the push for privatization. Air India was acquired by the Tata Group
in 2022, aiming to rejuvenate the carrier through reforms and fleet
revamping. PIA and SriLankan Airlines also consider privatization to
address their financial woes. The challenges include political
interference, debt, and management inefficiencies. Privatization
offers benefits like professional management and operational freedom
but requires government support to succeed. The path to privatization
in South Asia faces shared challenges and the need for long-term
commitments to restructuring.

import openai
import pandas as pd

# Set your OpenAI API key

openai.api_key = "[redacted]"

# Summarization function tailored to your project
def summarize_article(article_text):
    try:
        response = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system", "content": "You are an expert
summarizer for articles related to aviation and privatization."},
                {"role": "user", "content": f"Summarize the following
article in 100 words or less: {article_text}"},
            ],
            max_tokens=150,  # Limit response length
            temperature=0.5  # Make the output deterministic
        )
        return response['choices'][0]['message']['content'].strip()
    except Exception as e:
        return f"Error: {str(e)}"
```

```python
# Load the dataset
file_path = "cleaned_air_india_case_study_articles.csv"  # Adjust to
your file path
df = pd.read_csv(file_path)

# Ensure 'Content' column exists and is not empty
if 'Content' not in df.columns or df['Content'].isnull().all():
    raise ValueError("The 'Content' column is missing or empty in the
dataset.")

# Drop rows with missing content
df = df.dropna(subset=['Content']).reset_index(drop=True)

# Apply summarization to all articles
df['Summary'] = df['Content'].apply(summarize_article)

# Display a few rows with the original content and the generated
summary
print("Preview of Summarization Results:")
print(df[['Content', 'Summary']].head())  # Adjust number of rows with
.head(n) as needed

# Save the updated dataset with summaries
output_file_path = "summarized_air_india_data.csv"
df.to_csv(output_file_path, index=False)

print(f"Summarization completed! Results saved to
'{output_file_path}'.")
```

```
Preview of Summarization Results:
                                             Content  \
0  Commercial Aviation News and Analysis india fl...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

                                             Summary
0  The article discusses the trend of privatizing...
1  Privatization of flag carriers in the Indian s...
2  On November 11, Vistara will merge with Air In...
3  The article covers a range of topics including...
4  Two years after Tata Group acquired Air India,...
Summarization completed! Results saved to
'summarized_air_india_data.csv'.
```

```python
# Classification Function
def classify_article(article_text):
    categories = ["Privatization Benefits", "Challenges", "Market
```

```
Growth", "Customer Experience"]
    prompt = f"Classify the following article into one of these
categories: {', '.join(categories)}:\n\n{article_text}"

    try:
        response = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system", "content": "You are an expert
classifier for aviation and privatization articles."},
                {"role": "user", "content": prompt},
            ],
            max_tokens=50,
            temperature=0.5  # Lower randomness for consistency
        )
        return response['choices'][0]['message']['content'].strip()
    except Exception as e:
        return f"Error: {str(e)}"

# Apply classification to all articles
df['Category'] = df['Content'].apply(classify_article)

# Save intermediate dataset with classification results
classification_output_path = 'classified_air_india_data.csv'
df.to_csv(classification_output_path, index=False)

# Preview results
print("Classification Completed! Preview:")
print(df[['Content', 'Category']].head())
```

```
Classification Completed! Preview:
                                             Content  \
0  Commercial Aviation News and Analysis india fl...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

                                          Category
0                 Category: Privatization Benefits
1                 Category: Privatization Benefits
2                         Category: Market Growth
3  This article does not fall into any of the pro...
4                              Category: Challenges
```

```
# Sentiment Analysis Function
def analyze_sentiment(article_text):
    prompt = f"Analyze the sentiment of the following article and
classify it as Positive, Negative, or Neutral:\n\n{article_text}"
```

```
    try:
        response = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system", "content": "You are an expert
sentiment analyzer for aviation-related articles."},
                {"role": "user", "content": prompt},
            ],
            max_tokens=50,
            temperature=0.5
        )
        return response['choices'][0]['message']['content'].strip()
    except Exception as e:
        return f"Error: {str(e)}"

# Apply sentiment analysis to all articles
df['Sentiment'] = df['Content'].apply(analyze_sentiment)

# Save intermediate dataset with sentiment analysis results
sentiment_output_path = 'sentiment_air_india_data.csv'
df.to_csv(sentiment_output_path, index=False)

# Preview results
print("Sentiment Analysis Completed! Preview:")
print(df[['Content', 'Sentiment']].head())
```

```
Sentiment Analysis Completed! Preview:
                                                Content  \
0  Commercial Aviation News and Analysis india fl...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

                                           Sentiment
0  The sentiment of the article is mostly Negativ...
1  The sentiment of the article is Neutral. The a...
2  The sentiment of the article is mostly Neutral...
3  The sentiment of the article appears to be Neu...
4  The sentiment of the article can be classified...
```

```
# Combine all results
df['Summary'] = df['Summary']  # From summarization step
df['Category'] = df['Category']  # From classification step
df['Sentiment'] = df['Sentiment']  # From sentiment analysis step

# Save final dataset
final_output_path = 'final_air_india_data.csv'
df.to_csv(final_output_path, index=False)
```

```python
# Preview final dataset
print("Final Dataset Preview:")
print(df[['Content', 'Summary', 'Category', 'Sentiment']].head())
```

```
Final Dataset Preview:
                                             Content  \
0  Commercial Aviation News and Analysis india fl...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

                                             Summary  \
0  The article discusses the trend of privatizing...
1  Privatization of flag carriers in the Indian s...
2  On November 11, Vistara will merge with Air In...
3  The article covers a range of topics including...
4  Two years after Tata Group acquired Air India,...

                                            Category  \
0                  Category: Privatization Benefits
1                  Category: Privatization Benefits
2                           Category: Market Growth
3  This article does not fall into any of the pro...
4                               Category: Challenges

                                           Sentiment
0  The sentiment of the article is mostly Negativ...
1  The sentiment of the article is Neutral. The a...
2  The sentiment of the article is mostly Neutral...
3  The sentiment of the article appears to be Neu...
4  The sentiment of the article can be classified...
```

```python
# Extract the main sentiment
df['Cleaned Sentiment'] = df['Sentiment'].apply(lambda x: x.split()
[4])

# Extract the main category
df['Cleaned Category'] = df['Category'].apply(lambda x: x.split(':')[-
1].strip())

# Drop the old columns if necessary
df = df.drop(columns=['Category', 'Sentiment'])

# Rename columns for clarity
df.rename(columns={'Cleaned Sentiment': 'Sentiment', 'Cleaned
Category': 'Category'}, inplace=True)

# Preview the cleaned dataset
```

```
print("Refined Dataset Preview:")
print(df[['Content', 'Summary', 'Category', 'Sentiment']].head())
```

```
Refined Dataset Preview:
                                             Content  \
0  Commercial Aviation News and Analysis india fl...
1  In recent years, the trend of privatizing flag...
2  Subscribe today to keep up with the latest tra...
3  How to make vrat-friendly Coconut Peanut Chutn...
4  Two years after the Tata Group's acquisition o...

                                             Summary  \
0  The article discusses the trend of privatizing...
1  Privatization of flag carriers in the Indian s...
2  On November 11, Vistara will merge with Air In...
3  The article covers a range of topics including...
4  Two years after Tata Group acquired Air India,...

                                        Category Sentiment
0                     Privatization Benefits   article
1                     Privatization Benefits   article
2                              Market Growth   article
3  This article does not fall into any of the pro...   article
4                                 Challenges   article
```

```
# Combine relevant articles about challenges before privatization
pre_articles = "\n\n".join(df[df['Category'] == 'Privatization
Benefits']['Content'].head(3))

# Prompt for pre-privatization analysis
prompt_pre = f"""
Analyze the following news articles about Air India before its
privatization:
{pre_articles}

Answer the following:
1. What were the main operational and financial challenges Air India
faced before privatization?
2. How did political interference contribute to Air India’s struggles?
3. Why were earlier government bailouts insufficient to save Air
India?
Provide a detailed summary based on the articles.
"""

# Call LLM
response_pre = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation industry
expert."},
```

```
        {"role": "user", "content": prompt_pre}
    ],
    max_tokens=400,
    temperature=0.7
)

# Print results
print("Pre-Privatization Challenges Analysis:")
print(response_pre['choices'][0]['message']['content'])
```

```
Pre-Privatization Challenges Analysis:
Certainly, I will analyze the provided news articles about Air India before its privatization.

1. Main operational and financial challenges faced by Air India before privatization:
- The articles highlight that Air India was burdened with massive debt, estimated to be around $8 billion, and was consistently incurring losses for years.
- Operational inefficiencies, high operating costs, and fierce competition in the aviation industry were major challenges faced by the airline.
- The airline's aging fleet, labor disputes, and overstaffing were also noted as significant operational challenges that impacted its performance and cost structure.

2. Political interference and its contribution to Air India's struggles:
- Political interference, as per the articles, played a detrimental role in Air India's struggles. Decisions on fleet acquisition, route planning, and appointments of top management were often influenced by political considerations rather than commercial viability.
- The airline was reportedly used by politicians for their travel needs, leading to preferential treatment and inefficiencies in operations.
- Political interference also hindered the implementation of necessary reforms and restructuring initiatives that could have potentially improved Air India's financial health and operational efficiency.

3. Inadequacy of earlier government bailouts to save Air India:
- The news articles suggest that previous government bailouts provided to Air India were insufficient to address the core issues plaguing the airline.
- The bailout funds were primarily used to service the airline's mounting debt and cover operational losses rather than investing in strategic initiatives to enhance competitiveness and sustainability.
- Lack of stringent conditions and oversight in the bailout packages also allowed Air India to continue with its inefficient practices without implementing substantial reforms.
```

```
Summary:
Before its privatization, Air India grappled with a myriad of
challenges, including massive debt, operational inefficiencies, and
political interference. The airline's financial woes were exacerbated
by its aging fleet, labor disputes, and intense competition in the
industry. Political interference in decision-making processes hindered
the airline's ability to make strategic and commercially viable
choices, leading to further operational struggles
```

```python
# Combine relevant articles about improvements after privatization
post_articles = "\n\n".join(df[df['Category'] == 'Market Growth']
['Content'].head(3))

# Prompt for post-privatization analysis
prompt_post = f"""
Analyze the following news articles about Air India after its
privatization:
{post_articles}

Answer the following:
1. What leadership and operational changes has the Tata Group
implemented since 2022?
2. How has fleet modernization contributed to Air India’s recovery?
3. What has been the customer and market response to Air India’s
privatization?
Provide a detailed summary based on the articles.
"""

# Call LLM
response_post = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation analyst
specializing in privatization."},
        {"role": "user", "content": prompt_post}
    ],
    max_tokens=400,
    temperature=0.7
)

# Print results
print("Post-Privatization Improvements Analysis:")
print(response_post['choices'][0]['message']['content'])
```

```
Post-Privatization Improvements Analysis:
I'm sorry, but I cannot provide a detailed summary based on specific
news articles since I don't have access to real-time information.
However, I can offer a general analysis of how privatization may
impact Air India based on trends and past experiences in the aviation
industry.
```

1. Leadership and Operational Changes: After privatization by the Tata Group, significant leadership and operational changes are expected to take place. The Tata Group is known for its efficient and strategic management style, so one can anticipate a focus on improving operational efficiency, enhancing customer service, and cost-cutting measures. There may be restructuring of the management team, streamlining of processes, and implementation of new technologies to optimize operations.

2. Fleet Modernization: Fleet modernization is crucial for any airline's recovery and growth. With privatization, Air India may have access to better financial resources to invest in upgrading and expanding its fleet. Newer and more fuel-efficient aircraft can help reduce operational costs, improve reliability, and enhance the overall customer experience. This can also open up new routes and markets for Air India, leading to increased revenue and market share.

3. Customer and Market Response: Privatization of Air India may lead to a more customer-centric approach, focusing on improving service quality, on-time performance, and overall passenger experience. The market response to privatization will depend on how effectively the Tata Group can implement changes to make Air India more competitive and attractive to passengers. Positive changes in pricing, routes, and services can help regain customer trust and loyalty, leading to an increase in market share.

Overall, privatization of Air India by the Tata Group is likely to bring about significant changes in leadership, operations, fleet modernization, and customer service. These changes are essential for the airline's recovery and long-term success in a highly competitive aviation industry.

```
# Combine articles about financial and operational performance
financial_articles = "\n\n".join(df[df['Category'] == 'Challenges']
['Content'].head(3))

# Prompt for financial analysis
prompt_financial = f"""
Analyze the following news articles discussing Air India's financial
and operational performance:
{financial_articles}

Answer the following:
1. How has Air India's revenue, profitability, and operational
efficiency evolved pre- and post-privatization?
2. What are the key investments made in fleet, technology, and
infrastructure?
3. How do these investments align with the improvements in
performance?
```

```
"""

# Call LLM
response_financial = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation financial
analyst."},
        {"role": "user", "content": prompt_financial}
    ],
    max_tokens=400,
    temperature=0.7
)

# Print results
print("Financial and Operational Performance Analysis:")
print(response_financial['choices'][0]['message']['content'])
```

```
Financial and Operational Performance Analysis:
I'm sorry, but I currently don't have access to real-time news
articles or specific data on Air India's financial and operational
performance to provide a detailed analysis. However, I can offer a
general approach on how to analyze the information once it's
available.

1. Revenue, Profitability, and Operational Efficiency Pre- and Post-
Privatization:

- Analyze Air India's financial statements to compare revenue trends,
profit margins, and operational efficiency metrics (such as load
factor, on-time performance) before and after privatization.
- Look for any significant changes in revenue sources, cost structure,
and profitability ratios.
- Consider factors like competition, market dynamics, regulatory
environment, and management decisions that may have influenced the
airline's performance.

2. Key Investments in Fleet, Technology, and Infrastructure:

- Review Air India's annual reports or press releases to identify
major investments in fleet expansion, modernization, and fleet
renewal.
- Evaluate the adoption of new technologies for operational
efficiency, customer service enhancements, and cost optimization.
- Assess investments in infrastructure improvements at airports,
maintenance facilities, and IT systems to support growth and enhance
service quality.

3. Alignment of Investments with Performance Improvements:
```

```
- Connect the timeline of investments with changes in key performance indicators like revenue growth, cost reduction, customer satisfaction, and operational reliability.
- Determine how fleet upgrades or technology investments have contributed to increased revenue generation, cost savings, or improved customer experience.
- Evaluate the impact of infrastructure investments on operational efficiency, route expansion, and overall competitiveness in the market.

By conducting a comprehensive analysis of Air India's financial and operational data alongside information on investments and performance, you can provide insights into the airline's strategic decisions and their outcomes on its business performance.

# Combine articles about comparisons with other airlines
comparison_articles = "\n\n".join(df[df['Category'] == 'Comparative Analysis']['Content'].head(3))

# Prompt for comparative analysis
prompt_comparison = f"""
Analyze the following news articles comparing Air India with other flag carriers like PIA and SriLankan Airlines:
{comparison_articles}

Answer the following:
1. How does Air India's privatization approach differ from PIA and SriLankan Airlines?
2. What lessons can other struggling airlines learn from Air India's journey?
Provide a concise comparison based on the articles.
"""

# Call LLM
response_comparison = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an expert in aviation comparative analysis."},
        {"role": "user", "content": prompt_comparison}
    ],
    max_tokens=400,
    temperature=0.7
)

# Print results
print("Comparative Analysis:")
print(response_comparison['choices'][0]['message']['content'])
```

```
Comparative Analysis:
Based on the news articles comparing Air India with PIA and SriLankan
Airlines, here is a concise comparison:

1. Privatization Approach:
- Air India: Air India's privatization approach involved a competitive
bidding process, with the government divesting its stake in the
airline. The focus was on finding a strategic buyer who could bring in
capital and expertise to turn around the airline's fortunes.
- PIA: PIA's privatization process has been marred by challenges, with
changes in government and lack of consensus on the best way forward.
The airline has faced resistance from labor unions and political
interference, delaying the privatization process.
- SriLankan Airlines: SriLankan Airlines has also struggled with
privatization efforts, facing issues with finding a suitable investor
and political instability affecting the process.

2. Lessons for Struggling Airlines:
- Transparency and Accountability: Air India's journey highlights the
importance of transparency and accountability in the privatization
process. Clear communication and a structured approach can help build
trust with potential investors and stakeholders.
- Strategic Partnerships: Struggling airlines can learn from Air
India's focus on finding a strategic buyer who can bring in not just
capital but also industry expertise and a long-term vision for the
airline's growth.
- Employee Engagement: Air India's experience also underscores the
importance of engaging with employee unions and stakeholders early on
in the privatization process to address concerns and build consensus
for the way forward.

In summary, while Air India's privatization approach focused on
transparency, strategic partnerships, and employee engagement, PIA and
SriLankan Airlines have faced challenges due to political
interference, lack of consensus, and difficulties in finding suitable
investors. Struggling airlines can learn from Air India's journey by
adopting a structured and transparent approach to privatization,
focusing on finding the right strategic partner, and engaging with
employees and stakeholders effectively.
```

```python
# Combine articles about Air India's future
future_articles = "\n\n".join(df[df['Category'] == 'Future Outlook']
['Content'].head(3))

# Prompt for future outlook
prompt_future = f"""
Analyze the following news articles discussing the future outlook of
Air India:
{future_articles}
```

```
Answer the following:
1. What are Air India's strategic initiatives for growth under Tata Group?
2. How does Air India plan to position itself in the Indian and global aviation markets?
3. What steps is Air India taking towards sustainability and reducing its carbon footprint?
"""

# Call LLM
response_future = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation strategy consultant."},
        {"role": "user", "content": prompt_future}
    ],
    max_tokens=400,
    temperature=0.7
)

# Print results
print("Future Outlook Analysis:")
print(response_future['choices'][0]['message']['content'])
```

```
Future Outlook Analysis:
I'm sorry, but I currently do not have access to external content such as news articles to provide a detailed analysis of Air India's strategic initiatives under Tata Group. However, I can provide you with some general insights based on typical strategic initiatives that airlines often undertake when looking to grow and position themselves in the market.

1. Strategic Initiatives for Growth under Tata Group:
a. Fleet Expansion: Air India may look to expand its fleet by acquiring new aircraft or modernizing its existing fleet to enhance operational efficiency and increase capacity.
b. Route Expansion: The airline could focus on expanding its domestic and international route network to connect more cities and cater to a broader customer base.
c. Service Enhancement: Improving customer service, inflight experience, and overall operational efficiency to enhance its competitive positioning in the market.
d. Alliances and Partnerships: Forming strategic alliances with other airlines or service providers to expand its reach and offer more seamless travel options to passengers.
e. Digital Transformation: Investing in digital technologies to enhance the booking process, improve customer engagement, and streamline operations.
```

```
2. Positioning in the Indian and Global Aviation Markets:
Air India may aim to position itself as a premium full-service carrier
catering to both domestic and international travelers. The airline
could focus on offering a mix of competitive pricing, quality service,
and a diverse route network to attract different customer segments. In
the global market, Air India may leverage its strategic partnerships
and alliances to strengthen its presence and increase market share.

3. Steps Towards Sustainability and Carbon Footprint Reduction:
a. Fleet Modernization: Investing in newer, more fuel-efficient
aircraft to reduce emissions per passenger-kilometer.
b. Sustainable Aviation Fuels: Exploring the use of sustainable
aviation fuels to reduce the carbon footprint of its operations.
c. Operational Efficiency: Implementing measures to optimize flight
operations, reduce fuel consumption, and minimize waste generation.
d. Carbon Offsetting: Participating in carbon offset programs to
compensate for its greenhouse gas emissions.
e. Green
```

```python
with open('air_india_llm_analysis.txt', 'w') as f:
    f.write("Pre-Privatization Challenges Analysis:\n")
    f.write(response_pre['choices'][0]['message']['content'] + "\n\n")
    f.write("Post-Privatization Improvements Analysis:\n")
    f.write(response_post['choices'][0]['message']['content'] + "\n\
n")
    f.write("Financial and Operational Performance Analysis:\n")
    f.write(response_financial['choices'][0]['message']['content'] +
"\n\n")
    f.write("Comparative Analysis:\n")
    f.write(response_comparison['choices'][0]['message']['content'] +
"\n\n")
    f.write("Future Outlook Analysis:\n")
    f.write(response_future['choices'][0]['message']['content'] + "\n\
n")

# Prompt for refining future outlook
selected_articles = "\n\n".join(df['Content'].head(5))  # Adjust as
necessary

prompt_refined = f"""
Using the following articles about Air India, provide a detailed
analysis of its future outlook under the Tata Group:
{selected_articles}

Specifically, address:
1. What strategic initiatives has Air India undertaken for growth and
modernization?
2. How does Air India plan to position itself in the Indian and global
markets post-privatization?
3. What steps is Air India taking toward sustainability and reducing
```

```
its carbon footprint?
"""

response_refined = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation strategy
expert."},
        {"role": "user", "content": prompt_refined}
    ],
    max_tokens=500,
    temperature=0.7
)

print("Refined Future Outlook Analysis:")
print(response_refined['choices'][0]['message']['content'])
```

```
Refined Future Outlook Analysis:
I'm currently unable to browse the internet to access the specific
articles mentioned. However, based on my expertise in aviation
strategy, I can provide a general analysis of Air India's future
outlook under the Tata Group.

1. Strategic Initiatives for Growth and Modernization:
Under the Tata Group's ownership, Air India is likely to embark on a
series of strategic initiatives aimed at driving growth and
modernization. These initiatives may include fleet renewal and
expansion, route optimization, improved customer experience, and
operational efficiency enhancement. Tata Group's deep industry
expertise and financial strength can potentially support Air India in
upgrading its fleet with newer and more fuel-efficient aircraft,
enhancing its network connectivity, and investing in advanced
technologies to streamline operations and improve overall performance.

2. Positioning in Indian and Global Markets Post-Privatization:
As Air India transitions into a privately-owned entity under the Tata
Group, it is expected to focus on repositioning itself in both the
Indian and global aviation markets. The airline may adopt a more
customer-centric approach, offering competitive pricing, enhanced
services, and improved connectivity to strengthen its market share
domestically. Internationally, Air India could leverage Tata Group's
global network and reputation to expand its presence, explore new
routes, and forge strategic partnerships with other airlines to
enhance its global reach and competitiveness.

3. Sustainability and Carbon Footprint Reduction:
In line with the aviation industry's increasing focus on
sustainability, Air India, under the Tata Group's ownership, is likely
to prioritize initiatives aimed at reducing its carbon footprint and
enhancing environmental sustainability. The airline may invest in more
```

```
fuel-efficient aircraft, adopt sustainable aviation fuels, implement
eco-friendly practices in its operations, and participate in carbon
offset programs to mitigate its environmental impact. By demonstrating
a commitment to sustainability, Air India can enhance its reputation,
attract environmentally-conscious passengers, and contribute to the
industry's efforts to combat climate change.

Overall, the future outlook for Air India under the Tata Group appears
promising, with opportunities for growth, modernization, market
positioning, and sustainability initiatives that can strengthen the
airline's competitiveness and long-term success in the dynamic
aviation industry.
```

```
# Prompt for summarizing key themes
all_articles = "\n\n".join(df['Content'].head(10))  # Adjust as
necessary

prompt_summary = f"""
Summarize the key themes across the following articles about Air
India:
{all_articles}

Group the themes into:
1. Pre-Privatization Challenges
2. Post-Privatization Improvements
3. Future Growth and Market Positioning
4. Sustainability Goals

Provide concise, theme-based summaries for each group.
"""

response_summary = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an expert in
synthesizing themes from articles."},
        {"role": "user", "content": prompt_summary}
    ],
    max_tokens=600,
    temperature=0.7
)

print("Key Themes Summary:")
print(response_summary['choices'][0]['message']['content'])
```

```
Key Themes Summary:
**1. Pre-Privatization Challenges:**
- High debt burden: Air India has been struggling with a significant
debt burden, leading to financial instability and operational
challenges.
```

```
- Inefficiencies and bureaucracy: The airline has faced issues related to inefficiencies, bureaucratic hurdles, and lack of flexibility in decision-making processes.
- Competition and market share erosion: Air India's market share has been declining due to fierce competition from other airlines, impacting its overall performance and profitability.


**2. Post-Privatization Improvements:**
- Operational efficiency: Privatization has brought in a renewed focus on operational efficiency, cost-cutting measures, and improved management practices.
- Customer service enhancements: The airline has made efforts to enhance customer service, improve passenger experience, and introduce new services to stay competitive in the market.
- Fleet modernization: Post-privatization, Air India has invested in fleet modernization, introducing new aircraft and technologies to enhance its services and improve operational capabilities.


**3. Future Growth and Market Positioning:**
- Expansion and route optimization: Air India is looking to expand its route network, explore new markets, and optimize its route structure to enhance its market positioning and competitiveness.
- Strategic partnerships and alliances: The airline is seeking strategic partnerships and alliances with other airlines to strengthen its market position, increase connectivity, and tap into new customer segments.
- Innovation and digital transformation: Air India is focusing on innovation and digital transformation initiatives to adapt to changing market dynamics, improve efficiency, and enhance customer engagement.


**4. Sustainability Goals:**
- Environmental initiatives: Air India has set sustainability goals to reduce its carbon footprint, increase energy efficiency, and adopt environmentally friendly practices to contribute to environmental conservation.
- Social responsibility: The airline is committed to social responsibility initiatives, including community engagement, employee welfare programs, and support for underprivileged communities.
- Financial sustainability: Air India aims to achieve long-term financial sustainability through strategic planning, cost management, revenue optimization, and investment in sustainable business practices.

# Prompt for lessons other airlines can learn
prompt_lessons = """
Based on Air India’s privatization journey, what lessons can be drawn for other struggling national airlines like PIA and SriLankan
```

```
Airlines? Address:
1. Operational improvements.
2. Financial restructuring.
3. Customer satisfaction enhancements.
4. Sustainability initiatives.
"""

response_lessons = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation expert
providing lessons for airlines."},
        {"role": "user", "content": prompt_lessons}
    ],
    max_tokens=500,
    temperature=0.7
)

print("Lessons for Other Airlines:")
print(response_lessons['choices'][0]['message']['content'])
```

```
Lessons for Other Airlines:
Lessons from Air India's privatization journey for struggling national
airlines like PIA and SriLankan Airlines:

1. Operational Improvements:
- Focus on streamlining operations, reducing inefficiencies, and
enhancing overall operational performance.
- Implement modern technology, automation, and best practices to
improve efficiency and reliability.
- Optimize routes, fleet utilization, and crew scheduling to maximize
productivity and minimize costs.

2. Financial Restructuring:
- Conduct a thorough financial assessment to identify areas of cost
reduction and revenue enhancement.
- Implement strict cost control measures, including renegotiating
contracts, optimizing fuel consumption, and reducing overhead
expenses.
- Seek strategic partnerships or alliances to improve financial
stability and access to capital.

3. Customer Satisfaction Enhancements:
- Prioritize customer experience by investing in staff training,
service quality improvements, and customer engagement initiatives.
- Implement feedback mechanisms to gather customer insights and
address pain points proactively.
- Enhance loyalty programs, personalized services, and in-flight
amenities to differentiate from competitors and build customer
loyalty.
```

```
4. Sustainability Initiatives:
- Implement environmentally friendly practices to reduce carbon
footprint and comply with sustainability standards.
- Invest in fuel-efficient aircraft, alternative fuel sources, and
carbon offset programs to mitigate environmental impact.
- Engage in community outreach, corporate social responsibility
programs, and stakeholder engagement to demonstrate commitment to
sustainability.

Overall, a comprehensive approach that addresses operational,
financial, customer, and sustainability aspects is crucial for the
successful turnaround of struggling national airlines like PIA and
SriLankan Airlines. By learning from the experiences of Air India and
other successful privatization cases, these airlines can chart a path
towards long-term viability and competitiveness in the global aviation
market.
```

```python
# Prompt for strategic recommendations
prompt_recommendations = """
Based on Air India's current trajectory, provide strategic
recommendations for:
1. Enhancing operational efficiency and customer satisfaction.
2. Expanding its market share in India and globally.
3. Achieving sustainability goals while maintaining profitability.

Include actionable steps and potential challenges.
"""

response_recommendations = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are a strategy consultant
for the aviation industry."},
        {"role": "user", "content": prompt_recommendations}
    ],
    max_tokens=500,
    temperature=0.7
)

print("Strategic Recommendations:")
print(response_recommendations['choices'][0]['message']['content'])
```

```
Strategic Recommendations:
**1. Enhancing Operational Efficiency and Customer Satisfaction:**

**Recommendations:**
- Implement a comprehensive digital transformation strategy to
streamline operations, improve employee productivity, and enhance the
overall customer experience.
```

- Invest in advanced technology solutions such as AI-driven predictive maintenance, data analytics for route optimization, and automated customer service platforms.
- Enhance training programs for staff to improve service quality and efficiency.
- Implement a proactive maintenance schedule to minimize disruptions and delays.
- Develop a robust feedback mechanism to gather customer insights and continuously improve service offerings.

Actionable Steps:
- Conduct a thorough assessment of current operational processes and identify areas for improvement.
- Partner with industry experts to implement cutting-edge technology solutions.
- Provide regular training and upskilling opportunities for employees.
- Establish key performance indicators (KPIs) to measure operational efficiency and customer satisfaction.

Potential Challenges:
- Resistance to change from employees accustomed to traditional processes.
- Initial high costs associated with technology implementation.
- Integration challenges when incorporating new systems with existing infrastructure.

2. Expanding Market Share in India and Globally:

Recommendations:
- Focus on expanding domestic routes to tap into the growing demand for air travel within India.
- Forge strategic partnerships with other airlines to expand international connectivity.
- Enhance the frequent flyer program to attract and retain loyal customers.
- Develop targeted marketing campaigns to promote Air India's unique value proposition.
- Explore opportunities to enter new markets with high growth potential.

Actionable Steps:
- Conduct market research to identify key growth areas and customer segments.
- Negotiate codeshare agreements with partner airlines to expand route networks.
- Launch targeted marketing campaigns through digital channels and traditional media.
- Enhance customer engagement through personalized offers and benefits.

```
**Potential Challenges:**
- Intense competition in the aviation industry both domestically and internationally.
- Regulatory hurdles in entering new markets.
- Economic uncertainties impacting travel demand.

**3. Achieving Sustainability Goals while Maintaining Profitability:**

**Recommendations:**
- Invest in fuel-efficient aircraft and adopt sustainable aviation fuel (SAF) to reduce carbon emissions.
- Implement waste reduction programs and promote recycling initiatives.
- Develop partnerships with environmental organizations to support sustainability initiatives.
- Engage in community outreach programs to raise awareness about environmental conservation.
- Integrate sustainability practices across all operations to create a culture of environmental responsibility.

**Actionable Steps:**
- Conduct an environmental impact assessment to identify areas for improvement.
- Collaborate with suppliers to source eco-friendly
```

```python
# Prompt for future scenarios
prompt_scenarios = """
Analyze potential future scenarios for Air India under the Tata Group:
1. Best-case scenario: What does success look like for Air India in the next 5–10 years?
2. Worst-case scenario: What risks could derail its progress?
3. Likely scenario: What is the most realistic trajectory based on its current strategies?

Provide reasoning for each scenario.
"""

response_scenarios = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an aviation strategist analyzing future scenarios."},
        {"role": "user", "content": prompt_scenarios}
    ],
    max_tokens=600,
    temperature=0.7
)

print("Future Scenarios Analysis:")
print(response_scenarios['choices'][0]['message']['content'])
```

```
Future Scenarios Analysis:
1. Best-case scenario: In the best-case scenario, Air India under the Tata Group successfully undergoes a comprehensive restructuring and modernization program. This includes streamlining operations, improving efficiency, upgrading the fleet with new and fuel-efficient aircraft, enhancing customer service, and expanding routes to key international destinations. The airline leverages the strong brand reputation and financial backing of the Tata Group to become a leading player in the Indian aviation market. By focusing on innovation and customer experience, Air India captures a significant market share and achieves sustainable profitability within the next 5-10 years.

2. Worst-case scenario: The worst-case scenario for Air India under the Tata Group involves challenges such as intense competition, economic downturns, high fuel prices, and geopolitical instability. If the airline fails to effectively address these issues and implement necessary reforms, it may struggle to remain competitive in the market. Poor financial performance, operational inefficiencies, labor disputes, and regulatory hurdles could derail its progress and lead to continued losses or even bankruptcy. Failure to adapt to changing market dynamics and consumer preferences could also result in declining market share and reputation.

3. Likely scenario: The most likely scenario for Air India under the Tata Group lies somewhere between the best and worst-case scenarios. With the Tata Group's experience and resources, Air India is expected to make significant improvements in its operations, customer service, and financial performance. However, challenges such as regulatory hurdles, industry competition, and economic uncertainties may hinder its growth trajectory. The airline is likely to focus on optimizing routes, fleet modernization, cost management, and enhancing the overall passenger experience to remain competitive in the market. While profitability may not be achieved immediately, a gradual turnaround and improved market position are expected over the next 5-10 years.

Overall, the success of Air India under the Tata Group will depend on its ability to adapt to market changes, innovate, and execute a well-defined strategy that addresses key operational and financial challenges in the aviation industry.
```

```python
import openai

# Combine articles into one input (if they are summarized or can be concatenated)
all_articles = "\n\n".join(df['Processed Content'].head(10))  # Adjust to include relevant articles

# Define the prompt
prompt_conclusion = f"""
```

```
You are a professional analyst and writer tasked with drafting a
conclusion for a paper on Air India’s privatization.

Based on the following articles:
{all_articles}

Write a conclusion that:
1. Summarizes the key insights derived from these articles.
2. Discusses metrics such as fleet expansion, customer satisfaction
trends, and sustainability goals.
3. Predicts Air India's future trajectory, including its potential
global market positioning and competitive strengths.
4. Highlights broader implications for the aviation industry,
especially for other national carriers considering privatization.

Include actionable insights and recommendations wherever possible.
"""

# Generate the conclusion using OpenAI's LLM
response_conclusion = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an expert in aviation
strategy and professional writing."},
        {"role": "user", "content": prompt_conclusion}
    ],
    max_tokens=600,
    temperature=0.7
)

# Print the conclusion
conclusion = response_conclusion['choices'][0]['message']['content']
print("Generated Conclusion:\n")
print(conclusion)
```

```
Generated Conclusion:

Conclusion:

The articles analyzed provide valuable insights into the complex
process of Air India's privatization and its implications for the
aviation industry. Key takeaways include the challenges faced by the
airline in terms of financial stability, operational efficiency, and
market competitiveness. However, recent developments such as fleet
expansion, improvements in customer satisfaction trends, and a renewed
focus on sustainability goals indicate a potential turnaround for Air
India under private ownership.

The metrics of fleet expansion underscore the airline's commitment to
modernizing its fleet and enhancing operational efficiency. This,
```

```
coupled with a positive trend in customer satisfaction scores,
signifies a growing focus on service quality and passenger experience.
Furthermore, Air India's sustainability goals align with global trends
towards environmental responsibility, positioning the airline as a
more socially responsible player in the industry.

Looking ahead, Air India's future trajectory post-privatization
appears promising, with the potential to strengthen its global market
positioning and competitive strengths. By leveraging its extensive
network, brand recognition, and strategic partnerships, Air India
could carve out a niche in key markets and enhance its competitiveness
against regional and international carriers. Embracing digitalization
and innovation in service delivery could further differentiate Air
India and drive customer loyalty.

The privatization of Air India also holds broader implications for the
aviation industry, especially for other national carriers considering
a similar transition. The success of Air India's privatization could
serve as a blueprint for other state-owned airlines seeking financial
viability and operational efficiency. Implementing transparent
governance structures, fostering a customer-centric approach, and
investing in sustainable practices could be key strategies for
national carriers looking to thrive in a competitive market landscape.

In conclusion, Air India's privatization journey presents
opportunities for growth and transformation in the aviation sector. By
focusing on fleet modernization, enhancing customer experience, and
pursuing sustainability goals, Air India can position itself as a
strong player in the global market. The airline's experience also
offers valuable lessons for other national carriers embarking on the
path to privatization, emphasizing the importance of strategic
planning, innovation, and adaptability in a dynamic industry
environment.

# Assuming all relevant responses from your screenshots are stored in
variables.
# Replace the variables with your actual LLM response variables.

with open('air_india_llm_analysis.txt', 'w') as f:
    # Write Pre-Privatization Challenges
    f.write("Pre-Privatization Challenges Analysis:\n")
    f.write(response_pre['choices'][0]['message']['content'] + "\n\n")

    # Write Post-Privatization Improvements
    f.write("Post-Privatization Improvements Analysis:\n")
    f.write(response_post['choices'][0]['message']['content'] + "\n\
n")

    # Write Financial and Operational Performance
    f.write("Financial and Operational Performance Analysis:\n")
```

```
    f.write(response_financial['choices'][0]['message']['content'] +
"\n\n")

    # Write Comparative Analysis
    f.write("Comparative Analysis:\n")
    f.write(response_comparison['choices'][0]['message']['content'] +
"\n\n")

    # Write Future Outlook
    f.write("Future Outlook Analysis:\n")
    f.write(response_future['choices'][0]['message']['content'] + "\n\
n")

    # Add Future Scenarios
    f.write("Future Scenarios Analysis:\n")
    f.write(response_scenarios['choices'][0]['message']['content'] +
"\n\n")

    # Add Strategic Recommendations
    f.write("Strategic Recommendations:\n")
    f.write(response_recommendations['choices'][0]['message']
['content'] + "\n\n")

    # Add Lessons for Other Airlines
    f.write("Lessons for Other Airlines:\n")
    f.write(response_lessons['choices'][0]['message']['content'] + "\
n\n")

    # Add Key Themes Summary
    f.write("Key Themes Summary:\n")
    f.write(response_summary['choices'][0]['message']['content'] + "\
n\n")

    # Add Conclusion
    f.write("Conclusion:\n")
    f.write(response_conclusion['choices'][0]['message']['content'] +
"\n\n")

print("All analyses have been written to
'air_india_llm_analysis.txt'.")
```

```
All analyses have been written to 'air_india_llm_analysis.txt'.
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
!ls -lh /content/
```

```
total 27M
-rw-r--r-- 1 root root 2.2M Nov 22 00:01
air_india_case_study_articles.csv
-rw-r--r-- 1 root root  24K Nov 22 04:46 air_india_llm_analysis.txt
-rw-r--r-- 1 root root 2.6M Nov 22 00:04 air_india_with_sentiments.csv
-rw-r--r-- 1 root root 2.6M Nov 22 00:04 air_india_with_topics.csv
-rw-r--r-- 1 root root 3.8M Nov 22 03:41 classified_air_india_data.csv
-rw-r--r-- 1 root root 3.7M Nov 22 00:01
cleaned_air_india_case_study_articles.csv
drwx------ 7 root root 4.0K Nov 22 05:38 drive
-rw-r--r-- 1 root root 3.9M Nov 22 03:44 final_air_india_data.csv
-rw-r--r-- 1 root root 292K Nov 22 00:04 keyphrase_comparison.csv
-rw-r--r-- 1 root root  395 Nov 22 00:02
pre_post_sentiment_comparison.csv
drwxr-xr-x 1 root root 4.0K Nov 22 04:56 sample_data
-rw-r--r-- 1 root root 3.9M Nov 22 03:43 sentiment_air_india_data.csv
-rw-r--r-- 1 root root 3.8M Nov 22 03:34 summarized_air_india_data.csv

!mkdir -p /content/drive/MyDrive/Colab_Runtime_Files

!cp -r /content/*.csv /content/*.txt
/content/drive/MyDrive/Colab_Runtime_Files/

!ls -lh /content/drive/MyDrive/Colab_Runtime_Files/

total 27M
-rw------- 1 root root 2.2M Nov 22 05:41
air_india_case_study_articles.csv
-rw------- 1 root root  24K Nov 22 05:41 air_india_llm_analysis.txt
-rw------- 1 root root 2.6M Nov 22 05:41 air_india_with_sentiments.csv
-rw------- 1 root root 2.6M Nov 22 05:41 air_india_with_topics.csv
-rw------- 1 root root 3.8M Nov 22 05:41 classified_air_india_data.csv
-rw------- 1 root root 3.7M Nov 22 05:41
cleaned_air_india_case_study_articles.csv
-rw------- 1 root root 3.9M Nov 22 05:41 final_air_india_data.csv
-rw------- 1 root root 292K Nov 22 05:41 keyphrase_comparison.csv
-rw------- 1 root root  395 Nov 22 05:41
pre_post_sentiment_comparison.csv
-rw------- 1 root root 3.9M Nov 22 05:41 sentiment_air_india_data.csv
-rw------- 1 root root 3.8M Nov 22 05:41 summarized_air_india_data.csv
```