

PROYECTO: FINANCIAL NEWS SUMMARY

OBTENCIÓN DE DATOS

Se utilizó un DATA SET de encabezados de noticias financieras proveniente de la plataforma KAGGLE que proviene de noticias de las plataformas CNCB, THE GUARDIAN y REUTERS.

Load Data

```
# CNBC HEADLINES
cnbc_df = pd.read_csv(NEWS_PATH_CNBC)

# GUARDIAN HEADLINES
guardian_df = pd.read_csv(NEWS_PATH_GUARDIAN)

# REUTERS HEADLINES
reuters_df = pd.read_csv(NEWS_PATH_REUTERS)
```

[47]

✓ 0.0s

Python

PROCESAMIENTO DE DATOS

Después se hizo un procesamiento y estandarización de las fechas de las noticias para tenerlas como llaves para poder agrupar encabezados de noticias.

Data Processing

CNBC News

```
# Create 'Date' column
cnbc_df[['Month', 'Year', 'Date']] = cnbc_df['Time'].apply(
    lambda x: pd.Series(cncb_date_processing(x))
)

# Drop Auxiliar columns
cnbc_df.drop(['Month', 'Year'], axis=1, inplace=True)

# Drop News without 'Date'
cnbc_df.dropna(subset=['Date'], inplace=True)
```

[48]

✓ 0.0s

Generate Code Markdown

Python

```
cnbc_df_sel = cnbc_df[['Headlines', 'Date']]
```

[49]

✓ 0.0s

Python

The Guardian News

```
guardian_df['Date'] = guardian_df['Time'].apply(date_processing)
```

[50]

✓ 0.0s

Python

```
guardian_df_sel = guardian_df[['Headlines', 'Date']]
```

[51]

✓ 0.0s

Python

Reuters News

```
reuters_df['Date'] = reuters_df['Time'].apply(date_processing)
```

[52]

✓ 0.0s

Python

```
reuters_df_sel = reuters_df[['Headlines', 'Date']]
```

[70]

✓ 0.0s

Python

```
targetDate_cncb = "July 2020"
```

```
targetDate_others = "Jul 2020"
```

[143]

✓ 0.0s

Python

```
cnbc_df_fill = cnbc_df_sel[cnbc_df_sel['Date'] == targetDate_cncb][['Headlines']]
```

[127]

✓ 0.0s

Python

```
cnbc_df_fill.head(5)
```

[124]

✓ 0.0s

Python

```
Headlines
```

```
guardian_df_fill = guardian_df_sel[guardian_df_sel['Date'] == targetDate_others][['Headlines']]
```

[124]

✓ 0.0s

Python

```
guardian_df_fill.head(5)
```

[125]

✓ 0.0s

Python

```
Headlines
```

```
reuters_df_fill = reuters_df_sel[reuters_df_sel['Date'] == targetDate_others][['Headlines']]
```

[73]

✓ 0.0s

Python

```
reuters_df_fill.head(5)
```

[74]

✓ 0.0s

Python

```
Headlines
```

```
0 Johnson is asking Santa for a Christmas recovery
1 'I now fear the worst': four grim tales of wor...
2 Five key areas South must tackle to serve up...
3 Covid-19 leaves firms 'stagnantly disappointed' L...
4 The Week in Partnership | Philip Baccard's 'ad...
```

Se agruparon todos los encabezados de las noticias en un DataFrame.

Después se unieron todas las filas del DataFrame en una sola cadena de texto, dejando un enter entre los encabezados.

```
Get Corpus

inDate_news_corpus = "vuv".join(inDate_news["headline"].astype(str))
[41] ✓ 0.0s Python

print(inDate_news_corpus)
[41] ✓ 0.0s Python

...
Jim Cramer: A better way to invest in the Covid-19 vaccine gold rush

Cramer's lightning round: I would own Teradyne

Cramer's week ahead: Big week for earnings, even bigger week for vaccines

IQ Capital CEO Keith Bliss says tech and healthcare will rally

Wall Street delivered the 'kind of pullback I've been waiting for,' Jim Cramer says

Cramer's lightning round: I would just stay Long Mox

Acaras CEO: Parents can turn $5 into five figures for their kids through 'power of compounding'

Dividend cuts may mean rethinking your retirement income strategy

StackX has authorized 1 million Jordan sneakers this year, CEO says

Biohaven Pharmaceuticals lands Khloe Kardashian as influencer of new migraine drug

Cramer's lightning round: I like Beyond Meat

Cramer: We desperately need another round of federal stimulus, despite vaccine progress

Cramer's lightning round: Buy more Where

...

Boris Johnson's 'new deal' speech: what is missing?

How does Boris Johnson's 'new deal' compare with Franklin D Roosevelt's?

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings.
```

GENERACIÓN DEL PROMPT

Con esta cadena de texto larga y la de la fecha se utilizó el siguiente *prompt* para consultar el *API* de *OpenAI*.

```
Prompt Generation

prompt = """
You're an economic analyst.
Here are some financial headlines from {targetDate_others}:

{inDate_news_corpus}

Responds the following in JSON:

- "date": the input date,
- "summary": the economic summary,
- "optimism": @100,
- "sentiment_score": -1 to +1,
- "uncertainty": @100,
- "risk": @100,
- "topics": (list of main topics),
- "market_polarity": "bullish" or "bearish"

IMPORTANT NOTE: Return only the JSON, without additional text, preambles, or """json.

[36] ✓ 0.0s Python
```

SALIDAS DE LA CONSULTA

Las métricas que se esperan obtener con el modelo GPT de OpenAI es:

Score de Optimismo

Es una puntuación de 0 a 100 al optimismo económico (0 = pesimismo, 100 = extremo).

Nivel de incertidumbre

Puntaje de 0 a 100 sobre cuánta “incertidumbre económica” reflejan los titulares (mencionan crisis, volatilidad, cambios bruscos).

Confianza del consumidor / inversión

Score subjetivo derivado de las noticias sobre consumo, empleo, crédito.

Riesgo económico

Mide si los titulares apuntan a riesgos como inflación, recesión, defaults, tensiones políticas. Se sintetiza en un rango de 0 a 100

Tópicos principales

Lista de las palabras clave más relevantes en los titulares (ej. "inflation", "FED", "oil", "employment").

Polaridad de mercados

Si los titulares hablan más de subidas (bullish) o caídas (bearish).

MODELO UTILIZADO Y EJEMPLO DE SALIDA

Por último, el modelo que se usó para hacer la consulta es el **gpt-4o-mini**, con la intención de ahorrar recursos en la consulta.

A continuación se muestra una salida:

```
{
  "date": "July 2020",
  "summary": "The economic landscape in July 2020 is marked by a mix of cautious optimism and significant uncertainty. While there are signs of recovery in certain sectors, such as retail and technology, concerns about rising unemployment, potential job cuts, and the impact of ongoing lockdown measures persist. The market is reacting to both positive developments, like vaccine progress, and negative news, such as the fallout from factory scandals and the potential for further economic stimulus.",
  "optimism": 60,
  "sentiment_score": 0.2,
  "uncertainty": 75,
  "risk": 70,
  "topics": [
    "Covid-19 impact",
    "economic recovery",
    "unemployment",
    "stimulus measures",
    "retail sector",
    "technology",
    "healthcare",
    "market trends"
  ],
  "market_polarity": "bearish"
}
```

