

Resumen ejecutivo

Hito 2 del proyecto de clasificación de reclamaciones de TikTok

ISSUE / PROBLEM

El equipo de datos de TikTok busca desarrollar un modelo de aprendizaje automático para ayudar en la clasificación de las reclamaciones de los envíos de los usuarios. Para empezar, el equipo de datos debe organizar el conjunto de datos sin procesar y prepararlo para futuros análisis exploratorios de datos.

RESPONSE

El equipo de datos realizó una investigación preliminar del conjunto de datos de clasificación de siniestros con el objetivo de conocer las relaciones importantes entre las variables. Dada la petición de una clasificación de las reclamaciones de los usuarios, el equipo de datos analizó los recuentos de reclamaciones y opiniones para comprender el recuento de cada tipo de contenido de video.

IMPACT

El impacto de este análisis preliminar será evidente en los próximos pasos. Para comprender el impacto de los videos de los usuarios, el equipo de datos identificó dos variables importantes a considerar. Las variables video_duration (en segundos) y video_view_count son factores importantes a tener en cuenta para futuros modelos de predicción.

UNDERSTANDING THE DATA

Después de revisar el conjunto de datos proporcionado, la variable claim_status pareció particularmente útil, dado el proyecto propuesto por el cliente. Las siguientes capturas de pantalla muestran puntos importantes de análisis necesarios para comprender la variable claim_status.

```
data['claim_status'].value_counts()
```

```
claim      9608
opinion    9476
Name: claim_status, dtype: int64
```

Note: Los recuentos de cada estado de reclamación están bastante equilibrados. Hay 9.608 reclamaciones y 9.476 opiniones.

TENDENCIA DE PARTICIPACION

El equipo de datos tuvo en cuenta la participación de los espectadores con cada video en las categorías de reclamación y opinión. Para comprender la participación de los espectadores, el equipo de datos tuvo en cuenta el recuento de vistas. El recuento de vistas medias y medianas muestra el impacto de cada categoría de video; En concreto, los recuentos de visualizaciones medias y medianas de ambas categorías muestran la asociación entre el contenido (afirmación u opinión) y las visualizaciones de video.

Reclamaciones:

```
Mean view count claims: 501029.4527477102
Median view count claims: 501555.0
```

Opinions:

```
Mean view count opinions: 4956.43224989447
Median view count opinions: 4953.0
```

KEY INSIGHTS

- Hay un equilibrio casi igual entre opiniones y afirmaciones. Con esta comprensión, podemos continuar con nuestro análisis futuro sabiendo que hay una cantidad bastante equilibrada de afirmaciones y opiniones para los videos incluidos dentro de este conjunto de datos.
- Una vez identificadas las variables clave y la investigación inicial del conjunto de datos de clasificación de siniestros, puede comenzar el proceso de análisis exploratorio de los datos.

Pie chart visualizes the comparison of the count of claims and opinions

Total Number of Claims versus Opinions

