**Informe de la exploración de datos EDA**

**Resumen de hallazgos**

1. **Patrones**:
   * Se identificaron patrones de transacciones fraudulentas en ciertas categorías y horarios específicos.
   * Los fraudes tienden a ocurrir más frecuentemente durante la noche y los fines de semana.
2. **Relaciones**:
   * Existe una correlación significativa entre el número de transacciones y la probabilidad de fraude.
   * Los comerciantes con un alto número de transacciones diarias presentan un mayor riesgo de fraude.
3. **Anomalías**:
   * Se detectaron valores atípicos en las columnas amt y merchant\_risk\_30\_day, algunos de los cuales están asociados con fraudes.
   * Se encontraron duplicados en la columna trans\_num, que deben ser revisados y eliminados si es necesario.

**Columnas del DataFrame**

* trans\_date
* trans\_time
* unix\_time
* amt
* customer\_num\_trans\_1\_day
* customer\_num\_trans\_7\_day
* customer\_num\_trans\_30\_day
* customer\_avg\_amout\_1\_day
* customer\_avg\_amount\_7\_day
* customer\_avg\_amount\_30\_day
* merchant\_num\_trans\_1\_day
* merchant\_num\_trans\_7\_day
* merchant\_num\_trans\_30\_day
* merchant\_risk\_1\_day
* merchant\_risk\_7\_day
* merchant\_risk\_30\_day
* merchant\_risk\_90\_day
* is\_fraud
* trans\_num
* category
* merchant
* gender
* city
* state
* city\_pop
* job
* dob
* age
* profile
* lat
* long
* merch\_lat
* merch\_long
* trans\_time\_is\_night
* trans\_time\_day
* trans\_date\_is\_weekend

**Visualizaciones**

Para mejorar la comprensión de los datos y los hallazgos, se pueden incluir las siguientes visualizaciones:

1. **Gráficos de barras**:
   * **Distribución de transacciones fraudulentas por categoría**:
   * Gráfico, Gráfico de barras, Histograma

     Descripción generada automáticamente**Número de transacciones fraudulentas por hora del día**:

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

1. **Gráficos de dispersión**:
   * **Relación entre el monto de la transacción (amt) y la probabilidad de fraude**:
   * Gráfico

     Descripción generada automáticamente**Relación entre el riesgo del comerciante (merchant\_risk\_30\_day) y la probabilidad de fraude**:

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

1. **Mapas de calor**:
   * **Correlación entre diferentes variables numéricas**:

Imagen que contiene Texto

Descripción generada automáticamente

**Pasos para seguir**

1. **Normalizacion**:
   * **Normalización y estandarización**: Asegurarse de que las variables numéricas estén en una escala comparable.
2. **Análisis posterior**:
   * **Modelización de datos**: Utilizar técnicas de modelización como regresión logística, árboles de decisión o redes neuronales para predecir la variable is\_fraud.
   * **Validación cruzada**: Implementar técnicas de validación cruzada para evaluar la robustez de los modelos.
   * **Optimización de hiperparámetros**: Ajustar los hiperparámetros de los modelos para mejorar su rendimiento.
   * **Evaluación de modelos**: Utilizar métricas de evaluación como precisión, recall, F1-score, y AUC-ROC para medir el desempeño de los modelos.

**Conclusiones**

* La identificación de patrones y relaciones en los datos puede ayudar a desarrollar modelos predictivos más precisos.
* La limpieza y preparación de los datos son pasos cruciales para garantizar la calidad del análisis.
* Las visualizaciones pueden proporcionar una comprensión más clara de los hallazgos y facilitar la comunicación de los resultados.