**Contexto del Proyecto**

* **Empresa:** DSMarket, anteriormente TradiStores.
* **Transformación Digital:** DSMarket está en una fase de transformación digital liderada por Michelle Huggins, la nueva Chief Digital Officer. El objetivo es usar datos para optimizar operaciones y mejorar decisiones empresariales.

**Tareas Principales**

1. **Análisis de Datos:**
   * Analizar las ventas y otros datos de Nueva York, Boston y Filadelfia.
   * Identificar tendencias y productos populares por ciudad y tienda.
   * Crear un servicio de BI para el seguimiento regular de resultados.
2. **Clustering:**
   * Agrupar productos que se comportan de manera similar.
   * Evaluar la agrupación de tiendas según su comportamiento.
3. **Predicción de Ventas:**
   * Desarrollar un nuevo enfoque para la predicción de ventas a nivel de tienda-producto.
   * Realizar predicciones de ventas para un período de 28 días.
4. **Abastecimiento de Tiendas:**
   * Aplicar modelos predictivos al caso de uso de abastecimiento de tiendas.
   * Diseñar una prueba piloto para demostrar las mejoras y generar confianza en la solución.

**Datos Disponibles**

* **daily\_calendar\_with\_events.csv:** Información diaria sobre el calendario y eventos.
* **item\_prices.csv:** Precios de los productos por tienda y semana.
* **item\_sales.csv:** Ventas diarias de productos por tienda.

**Resultados Esperados**

* Documento técnico con la metodología y resultados.
* Entregables para cada tarea (dashboards, códigos, resultados de modelado).
* Predicciones de ventas para el conjunto de pruebas final.
* Presentación final para la junta ejecutiva.

**Evaluación**

* **Proyecto & memoria (70%):** 16/10/24
* **Presentación (30%):** 22/10/24

**Porcentajes de Evaluación**

* Análisis: 20%
* Clustering: 20%
* Modelo de predicción: 30%
* Caso de uso de abastecimiento de tiendas (con MLOps): 20%
* Diseño de la evaluación con piloto: 10%

**Próximos Pasos**

1. Cargar y explorar los datos proporcionados.
2. Definir el enfoque metodológico para cada tarea.
3. Asignar responsabilidades dentro del grupo de trabajo.
4. Crear un plan de trabajo detallado con hitos y fechas límite.

**Personajes Clave**

* **Michelle Huggins:** Chief Digital Officer con 15 años en Marketing Digital.
* **Nichole:** Científica de datos senior.
* **Paul Rogers:** CFO, responsable directo de Nichole.
* **Justin:** Recursos Humanos.
* **Martin:** Technology Lead.

**Expectativas del Proyecto: Primer Año**

* Estandarización y transformación de las fuentes de datos.
* Migración de todas las fuentes y procesos de datos a la nube.

**Prioridades del Departamento Financiero**

* **Predicciones de Ventas:** Minimizar margen de error debido a que los errores actuales son inaceptables.
* **Optimización con AI:** Procesos internos incluyendo estimaciones de stock, optimización de precios, entregas y predicciones de stockout (parte del plan a 5 años).

**Información y Datos Disponibles**

* **Ciudades:** Nueva York, Boston y Filadelfia.
* **Datos Disponibles:**
  + **daily\_calendar\_with\_events:**
    - **date:** día en formato año-mes-día.
    - **weekday:** día de la semana.
    - **weekday\_int:** día de la semana en número (Sábado=1, Viernes=7).
    - **d:** identificador de día.
    - **event:** nombre del evento (si existe).
  + **item\_prices:**
    - **item:** ID de los productos.
    - **category:** Categoría de los productos.
    - **store\_code:** Código alfanumérico de las tiendas.
    - **yearweek:** fecha del precio en formato año-semana.
    - **sell\_price:** precio promedio semanal del producto.
  + **item\_sales:**
    - **id:** ID de las ventas (combinación de item y store\_code).
    - **item:** ID de los productos.
    - **category:** Categoría de los productos.
    - **department:** ID del departamento.
    - **store:** Nombre de la tienda.
    - **store\_code:** Código alfanumérico de las tiendas.
    - **region:** Región.
    - **d\_1, d\_2, etc.:** número de unidades vendidas por día.

**Análisis Preliminar de los Datos**

**Daily Calendar with Events - 10 Filas Aleatorias:**

| **date** | **weekday** | **weekday\_int** | **d** | **event** |
| --- | --- | --- | --- | --- |
| 2015-07-24 | Friday | 7 | d\_1638 | NaN |
| 2013-05-06 | Monday | 3 | d\_829 | NaN |
| 2012-09-16 | Sunday | 2 | d\_597 | NaN |
| 2011-09-26 | Monday | 3 | d\_241 | NaN |
| 2011-06-13 | Monday | 3 | d\_136 | NaN |
| 2014-07-16 | Wednesday | 5 | d\_1265 | NaN |
| 2012-05-08 | Tuesday | 4 | d\_466 | NaN |
| 2015-08-31 | Monday | 3 | d\_1676 | NaN |
| 2014-04-27 | Sunday | 2 | d\_1185 | NaN |
| 2011-04-16 | Saturday | 1 | d\_78 | NaN |

**Valores Únicos:**

* **date:** 1913
* **weekday:** 7
* **weekday\_int:** 7
* **d:** 1913
* **event:** 5

**Item Prices - 10 Filas Aleatorias:**

| **item** | **category** | **store\_code** | **yearweek** | **sell\_price** |
| --- | --- | --- | --- | --- |
| ACCESORIES\_1\_306 | ACCESORIES | PHI\_3 | 201411.0 | 18.5934 |
| HOME\_&\_GARDEN\_2\_136 | HOME\_&\_GARDEN | BOS\_3 | 201217.0 | 6.0875 |
| ACCESORIES\_2\_037 | ACCESORIES | BOS\_3 | 201305.0 | 1.2901 |
| HOME\_&\_GARDEN\_1\_126 | HOME\_&\_GARDEN | BOS\_1 | 201122.0 | 9.5875 |
| HOME\_&\_GARDEN\_2\_299 | HOME\_&\_GARDEN | NYC\_2 | 201424.0 | 1.5500 |
| SUPERMARKET\_3\_060 | SUPERMARKET | NYC\_2 | 201545.0 | 3.5760 |
| ACCESORIES\_1\_015 | ACCESORIES | PHI\_3 | 201602.0 | 0.9310 |
| HOME\_&\_GARDEN\_1\_434 | HOME\_&\_GARDEN | PHI\_1 | 201604.0 | 4.9625 |
| SUPERMARKET\_3\_580 | SUPERMARKET | NYC\_1 | 201529.0 | 0.2400 |
| HOME\_&\_GARDEN\_1\_054 | HOME\_&\_GARDEN | NYC\_1 | 201443.0 | 4.9000 |

**Valores Únicos:**

* **item:** 3049
* **category:** 3
* **store\_code:** 10
* **yearweek:** 279
* **sell\_price:** 1892

**Item Sales - 10 Filas Aleatorias (continuación):**

| **id** | **item** | **category** | **department** | **store** | **store\_code** | **region** | **d\_1** | **d\_2** | **d\_3** | **...** | **d\_1904** | **d\_1905** | **d\_1906** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUPERMARKET\_1\_051\_BOS\_3 | SUPERMARKET\_1\_051 | SUPERMARKET | SUPERMARKET\_1 | Back\_Bay | BOS\_3 | Boston | 0 | 0 | 0 | ... | 0 | 1 | 0 |
| SUPERMARKET\_3\_779\_BOS\_2 | SUPERMARKET\_3\_779 | SUPERMARKET | SUPERMARKET\_3 | Roxbury | BOS\_2 | Boston | 0 | 0 | 0 | ... | 0 | 0 | 1 |
| ACCESORIES\_1\_387\_NYC\_4 | ACCESORIES\_1\_387 | ACCESORIES | ACCESORIES\_1 | Brooklyn | NYC\_4 | New York | 0 | 0 | 0 | ... | 25 | 0 | 2 |
| SUPERMARKET\_3\_115\_BOS\_1 | SUPERMARKET\_3\_115 | SUPERMARKET | SUPERMARKET\_3 | South\_End | BOS\_1 | Boston | 0 | 2 | 3 | ... | 0 | 1 | 1 |
| SUPERMARKET\_3\_289\_NYC\_3 | SUPERMARKET\_3\_289 | SUPERMARKET | SUPERMARKET\_3 | Tribeca | NYC\_3 | New York | 2 | 3 | 4 | ... | 2 | 2 | 1 |
| HOME\_&\_GARDEN\_2\_232\_PHI\_3 | HOME\_&\_GARDEN\_2\_232 | HOME\_&\_GARDEN | HOME\_&\_GARDEN\_2 | Queen\_Village | PHI\_3 | Philadelphia | 0 | 0 | 0 | ... | 1 | 1 | 2 |
| SUPERMARKET\_3\_574\_NYC\_3 | SUPERMARKET\_3\_574 | SUPERMARKET | SUPERMARKET\_3 | Tribeca | NYC\_3 | New York | 2 | 0 | 0 | ... | 6 | 9 | 7 |
| ACCESORIES\_1\_199\_PHI\_1 | ACCESORIES\_1\_199 | ACCESORIES | ACCESORIES\_1 | Midtown\_Village | PHI\_1 | Philadelphia | 0 | 0 | 0 | ... | 0 | 2 | 0 |
| ACCESORIES\_2\_112\_BOS\_1 | ACCESORIES\_2\_112 | ACCESORIES | ACCESORIES\_2 | South\_End | BOS\_1 | Boston | 0 | 0 | 0 | ... | 2 | 0 | 0 |
| HOME\_&\_GARDEN\_1\_366\_BOS\_1 | HOME\_&\_GARDEN\_1\_366 | HOME\_&\_GARDEN | HOME\_&\_GARDEN\_1 | South\_End | BOS\_1 | Boston | 0 | 0 | 0 | ... | 0 | 0 | 0 |

**Valores Únicos:**

* **id:** 30490
* **item:** 3049
* **category:** 3
* **department:** 7
* **store:** 10
* **store\_code:** 10
* **region:** 3
* **d\_1:** 2
* **d\_2:** 2
* **d\_3:** 3
* …
* **d\_1909:** 65
* **d\_1910:** 57
* **d\_1911:** 61
* **d\_1912:** 73
* **d\_1913:** 65

**Tareas Principales**

1. **Análisis (Marketing)**
   * **Requerimientos:**
     + Identificar productos que ya no son populares.
     + Identificar productos populares en diferentes ciudades y/o tiendas.
   * **Objetivo:** Dirigir acciones de marketing explotando estas diferencias.
   * **Entregable:** Dashboard en Power BI (20%).
2. **Clustering (Marketing)**
   * **Requerimientos:**
     + Identificar grupos de productos similares.
     + Determinar la cantidad de grupos.
     + Evaluar la similitud entre tiendas y la necesidad de su agrupación.
   * **Objetivo:** Evaluar el rendimiento de las campañas de marketing.
   * **Entregable:** Código y presentación de resultados (20%).
3. **Predicción de Ventas (Financiero)**
   * **Requerimientos:**
     + Actualizar el método actual de predicción tienda-producto.
     + Generar predicciones agregadas por departamento/tienda/ciudad.
   * **Objetivo:** Desarrollar un nuevo enfoque para la predicción de ventas a 28 días.
   * **Entregable:** Código y modelo predictivo (30%).
4. **Caso de Uso de Abastecimiento de Tiendas**
   * **Requerimientos:**
     + Aplicar modelos predictivos a la reposición en tiendas.
     + Minimizar el stock remanente semanal.
   * **Objetivo:** Proponer una solución detallada para aplicar los modelos de ventas al abastecimiento.
   * **Entregable:** Presentación de la propuesta (20%).
5. **Prueba Piloto**
   * **Requerimientos:**
     + Diseñar un estudio de casos y controles por productos o tiendas.
     + Evaluar y demostrar las mejoras.
   * **Objetivo:** Diseñar una prueba piloto para demostrar mejoras económicas.
   * **Entregable:** Presentación de la propuesta (10%).

**Plan de Acción (resumen)**

1. **Carga y Exploración de Datos:**
   * Importar los datos disponibles.
   * Realizar análisis exploratorios para entender la estructura y contenido.
2. **Definir Metodología:**
   * Establecer técnicas y herramientas específicas para cada tarea.
   * Asignar responsabilidades en el grupo de trabajo.
3. **Desarrollar y Validar Modelos:**
   * Implementar los modelos y algoritmos necesarios.
   * Validar los resultados con datos adicionales proporcionados.
4. **Preparar Entregables:**
   * Crear dashboards, presentaciones y documentos técnicos requeridos.
   * Ensayar la presentación final para la junta ejecutiva.

**Plan de Acción Detallado con Checkpoints y Checklists**

**1. Análisis de Datos (Marketing)**

**Objetivo: Dirigir acciones de marketing explotando diferencias en popularidad de productos.**

**Actividades:** 1.1. **Importación y limpieza de datos:**

* Importar datos de daily\_calendar\_with\_events.csv, item\_prices.csv, y item\_sales.csv.
* Limpiar y preparar los datos para análisis.

**Checklist:**

* Importar datos
* Limpiar datos (tratar valores nulos, duplicados)
* Preparar datos para análisis

**Checkpoint:**

* Reunión con el equipo para revisar el estado de la importación y limpieza de datos.
* Validación de los datos limpios y preparados.

1.2. **Análisis Exploratorio de Datos (EDA):**

* Realizar EDA para identificar tendencias y patrones en los datos.
* Utilizar visualizaciones para entender la distribución de ventas y precios por producto, tienda y ciudad.

**Checklist:**

* Realizar análisis descriptivo
* Crear visualizaciones iniciales
* Identificar patrones y tendencias clave

1.3. **Identificación de Productos Populares y No Populares:**

* Calcular métricas de ventas por producto y tienda.
* Identificar productos con ventas decrecientes y productos con ventas crecientes.

**Checklist:**

* Calcular métricas de ventas
* Identificar productos populares y no populares

**Checkpoint:**

* Reunión interna del equipo para revisar los productos identificados.
* Revisión de la metodología y resultados.

1.4. **Creación de Dashboards:**

* Desarrollar dashboards en Power BI para visualizar productos populares y no populares por ciudad y tienda.
* Incluir filtros y visualizaciones interactivas para facilitar el análisis.

**Checklist:**

* Diseñar estructura del dashboard
* Crear visualizaciones y filtros
* Integrar datos en Power BI

**Checkpoint:**

* Reunión con el tutor para presentar resultados preliminares del EDA.
* Reunión con el tutor para revisar el dashboard en Power BI.
* Feedback del tutor y ajustes finales.

**Entregables:**

* Dashboard en Power BI.
* Informe de análisis con hallazgos y recomendaciones.

**2. Clustering (Marketing)**

**Objetivo: Evaluar el rendimiento de las campañas de marketing agrupando productos y tiendas.**

**Actividades:** 2.1. **Preprocesamiento de Datos:**

* Normalizar y escalar los datos relevantes para clustering.
* Seleccionar variables clave para el clustering de productos y tiendas.

**Checklist:**

* Normalizar y escalar datos
* Seleccionar variables clave

**Checkpoint:**

* Reunión interna para revisar el preprocesamiento de datos.
* Validación de variables seleccionadas.

2.2. **Aplicación de Algoritmos de Clustering:**

* Utilizar algoritmos como K-means, DBSCAN o jerárquico para agrupar productos.
* Determinar el número óptimo de clusters utilizando métodos como el codo, la silueta y la validación cruzada.

**Checklist:**

* Aplicar K-means
* Aplicar DBSCAN
* Aplicar clustering jerárquico
* Determinar número óptimo de clusters

**Checkpoint:**

* Reunión con el equipo para revisar los resultados del clustering.
* Validación de clusters obtenidos.

2.3. **Análisis de Resultados:**

* Interpretar y visualizar los clusters resultantes.
* Evaluar la similitud entre tiendas y determinar la necesidad de agruparlas.

**Checklist:**

* Interpretar resultados
* Visualizar clusters
* Evaluar similitud entre tiendas

**Checkpoint:**

* Reunión con el tutor para presentar resultados del clustering.
* Feedback del tutor y ajustes necesarios.

2.4. **Informe de Resultados:**

* Documentar la metodología y los resultados del clustering.
* Proveer recomendaciones basadas en los clusters identificados.

**Checklist:**

* Documentar metodología
* Redactar informe de resultados
* Incluir recomendaciones

**Checkpoint:**

* Reunión interna para revisar el informe de resultados.
* Validación final del informe y recomendaciones.

**Entregables:**

* Código de clustering.
* Presentación de resultados con visualizaciones y recomendaciones.

**3. Predicción de Ventas (Financiero)**

**Objetivo: Desarrollar un nuevo enfoque para la predicción de ventas a 28 días.**

**Actividades:** 3.1. **Revisión y Mejoramiento del Método Actual:**

* Analizar el método actual de predicción de ventas.
* Identificar áreas de mejora y potenciales fuentes de error.

**Checklist:**

* Revisar método actual
* Identificar mejoras

**Checkpoint:**

* Reunión interna para discutir el método actual y posibles mejoras.
* Validación de áreas de mejora identificadas.

3.2. **Desarrollo del Modelo Predictivo:**

* Seleccionar algoritmos de predicción (e.g., ARIMA, LSTM, Prophet).
* Entrenar y validar el modelo utilizando datos históricos de ventas.

**Checklist:**

* Seleccionar algoritmos
* Entrenar modelo
* Validar modelo

**Checkpoint:**

* Reunión con el tutor para presentar el modelo desarrollado.
* Feedback del tutor y ajustes necesarios.

3.3. **Evaluación del Modelo:**

* Evaluar la precisión del modelo utilizando métricas como MAE, RMSE y MAPE.
* Comparar el rendimiento del nuevo modelo con el método actual.

**Checklist:**

* Evaluar precisión del modelo
* Comparar con el método actual

**Checkpoint:**

* Reunión interna para revisar la evaluación del modelo.
* Validación de comparaciones y ajustes finales.

3.4. **Implementación del Modelo:**

* Desplegar el modelo para realizar predicciones de ventas a 28 días.
* Automatizar la actualización del modelo con nuevos datos.

**Checklist:**

* Desplegar modelo
* Automatizar actualización

**Checkpoint:**

* Reunión con el tutor para revisar la implementación del modelo.
* Feedback del tutor y ajustes finales.

**Entregables:**

* Código del modelo predictivo.
* Informe con las predicciones y evaluación del modelo.

**4. Caso de Uso de Abastecimiento de Tiendas**

**Objetivo: Proponer una solución detallada para aplicar los modelos de ventas al abastecimiento.**

**Actividades:** 4.1. **Análisis de Requerimientos:**

* Entender el proceso actual de abastecimiento y sus desafíos.
* Definir los requerimientos específicos para el modelo de abastecimiento.

**Checklist:**

* Analizar proceso actual
* Definir requerimientos

**Checkpoint:**

* Reunión interna para discutir el proceso actual y requerimientos.
* Validación de requerimientos definidos.

4.2. **Desarrollo del Modelo de Abastecimiento:**

* Adaptar el modelo predictivo de ventas para el caso de uso de abastecimiento.
* Incorporar factores como el lead time, la capacidad de almacenamiento y las políticas de inventario.

**Checklist:**

* Adaptar modelo predictivo
* Incorporar factores adicionales

**Checkpoint:**

* Reunión con el tutor para presentar el modelo de abastecimiento desarrollado.
* Feedback del tutor y ajustes necesarios.

4.3. **Evaluación del Modelo:**

* Simular el desempeño del modelo de abastecimiento en diferentes escenarios.
* Evaluar la reducción del stock remanente y la mejora en la disponibilidad de productos.

**Checklist:**

* Simular desempeño
* Evaluar resultados

**Checkpoint:**

* Reunión interna para revisar la evaluación del modelo.
* Validación de simulaciones y resultados.

4.4. **Documentación de la Solución:**

* Redactar una propuesta detallada que incluya la metodología, resultados y recomendaciones para la implementación.
* Especificar las extensiones necesarias y los detalles de la productivización del modelo.

**Checklist:**

* Documentar metodología
* Redactar propuesta
* Incluir recomendaciones

**Checkpoint:**

* Reunión con el tutor para revisar la propuesta de solución.
* Feedback del tutor y ajustes finales.

**Entregables:**

* Presentación de la propuesta de solución.

**5. Prueba Piloto**

**Objetivo: Diseñar una prueba piloto para demostrar mejoras económicas.**

**Actividades:** 5.1. **Diseño de la Prueba Piloto:**

* Seleccionar productos y/o tiendas para el estudio de casos y controles.
* Definir métricas de evaluación y el diseño experimental.

**Checklist:**

* Seleccionar productos/tiendas
* Definir métricas de evaluación
* Diseñar experimento

**Checkpoint:**

* Reunión interna para discutir el diseño de la prueba piloto.
* Validación del diseño experimental y métricas definidas.

5.2. **Implementación de la Prueba Piloto:**

* Implementar la prueba piloto en las tiendas seleccionadas.
* Monitorear y recopilar datos durante el período de prueba.

**Checklist:**

* Implementar prueba piloto en tiendas seleccionadas
* Monitorear la ejecución de la prueba
* Recopilar datos durante el período de prueba

**Checkpoint:**

* Reunión interna para revisar el progreso de la prueba piloto.
* Ajustes necesarios en la implementación basada en el monitoreo inicial.

5.3. **Evaluación de Resultados:**

* Analizar los resultados de la prueba piloto para medir las mejoras económicas.
* Comparar el rendimiento de las tiendas y productos en la prueba piloto con los controles.

**Checklist:**

* Analizar resultados de la prueba piloto
* Comparar con grupos de control
* Medir mejoras económicas

**Checkpoint:**

* Reunión interna para revisar los resultados de la prueba piloto.
* Validación de análisis y comparaciones.

5.4. **Documentación y Presentación:**

* Documentar los hallazgos y las lecciones aprendidas de la prueba piloto.
* Preparar una presentación con los resultados y las recomendaciones para la expansión.

**Checklist:**

* Documentar hallazgos y lecciones aprendidas
* Preparar presentación de resultados
* Incluir recomendaciones para la expansión

**Checkpoint:**

* Reunión con el tutor para revisar la documentación y la presentación de la prueba piloto.
* Feedback del tutor y ajustes finales.

**Entregables:**

* Presentación de la propuesta de la prueba piloto.
* Informe de evaluación de la prueba piloto.

**Resumen de Checkpoints y Entregables**

1. **Análisis de Datos (Marketing):**
   * Checkpoints: Importación y limpieza de datos, EDA, identificación de productos, creación de dashboards.
   * Entregables: Dashboard en Power BI, informe de análisis.
2. **Clustering (Marketing):**
   * Checkpoints: Preprocesamiento de datos, aplicación de algoritmos, análisis de resultados, informe final.
   * Entregables: Código de clustering, presentación de resultados.
3. **Predicción de Ventas (Financiero):**
   * Checkpoints: Revisión del método actual, desarrollo del modelo, evaluación del modelo, implementación del modelo.
   * Entregables: Código del modelo predictivo, informe con predicciones y evaluación.
4. **Caso de Uso de Abastecimiento de Tiendas:**
   * Checkpoints: Análisis de requerimientos, desarrollo del modelo, evaluación del modelo, documentación de la solución.
   * Entregables: Presentación de la propuesta de solución.
5. **Prueba Piloto:**
   * Checkpoints: Diseño de la prueba piloto, implementación, evaluación de resultados, documentación y presentación.
   * Entregables: Presentación de la propuesta de la prueba piloto, informe de evaluación de la prueba piloto.