

Estudio de Caso: El Viaje de Lila para Convertirse en Científica de Datos: Su Enfoque de Trabajo en la Primera Tarea

Este estudio de caso explora la trayectoria profesional de un científico de datos y los atributos clave, destacando las habilidades, la educación y las experiencias necesarias para sobresalir en este campo dinámico. Seguiremos la historia de Lila, una persona ficticia que aspira a convertirse en una exitosa científica de datos.

Habrá un cuestionario después de esta lectura basado en el contenido de este estudio de caso.

Educación y Adquisición de Habilidades

Con un título de licenciatura en economía y una sólida experiencia en análisis de datos, Lila encuentra cautivadora la ciencia de datos y su potencial para impulsar un cambio significativo. Inspirada por sus experiencias, toma la decisión firme de cambiar su carrera y asumir el papel de científica de datos.

Lila se da cuenta de que para embarcarse en su viaje de ciencia de datos, necesita mejorar sus habilidades y conocimientos. Se inscribió en el programa en línea del Certificado Profesional en Ciencia de Datos de IBM, que cubre temas clave como estadísticas, aprendizaje automático, análisis de datos y lenguajes de programación como Python y SQL. Completa diligentemente el trabajo del curso y practica sus habilidades de codificación en conjuntos de datos reales.

Construyendo una Base Sólida

A medida que avanza en sus estudios, Lila adquiere una comprensión profunda de los fundamentos de la ciencia de datos, como la manipulación y visualización de datos con bibliotecas de Python como NumPy, Pandas y Matplotlib. Esta base sólida le proporciona habilidades esenciales para el análisis de datos.

Visualización para la Narración de Historias

Lila sabe que debe comunicar sus hallazgos de manera efectiva, por lo que aprende qué tipos de visualizaciones de datos serán más informativas. Aprende a crear gráficos y diagramas que representan visualmente datos como tendencias de ventas, segmentación de clientes y popularidad de productos, lo que permite a los interesados comprender la importancia de los datos. Estas visualizaciones ayudan en la narración de historias y en la toma de decisiones.

Experiencia Práctica

Lila entiende que la experiencia práctica es invaluable en la ciencia de datos. Comenzó a participar en competencias de Kaggle y a trabajar en proyectos de datos personales. Estas experiencias la exponen a problemas de datos del mundo real y le ayudan a desarrollar habilidades para resolver problemas. Además, creó su cuenta de GitHub y subió sus proyectos para construir su perfil.

Manipulación y Preprocesamiento de Datos

Lila aprende que los científicos de datos dedican una parte significativa de su tiempo a la limpieza y preprocesamiento de datos. Trabajó en varios conjuntos de datos, aprendió el preprocesamiento de datos mientras utilizaba las bibliotecas de Python NumPy y pandas, y se volvió experta en manejar datos faltantes, detección de valores atípicos y creación de características para mejorar el rendimiento del modelo.

Comunicación y Narración de Historias

Reconociendo que los científicos de datos deben comunicar sus hallazgos de manera efectiva, Lila perfeccionó sus habilidades de narración de datos. Aprendió diversas herramientas como matplotlib y plotly mientras cursaba su Certificado Profesional en Ciencia de Datos de IBM. Aprendió a crear visualizaciones atractivas y a presentar sus ideas de manera clara y comprensible.

Redes y Colaboración

Lila participa activamente en comunidades de ciencia de datos y asiste a encuentros y conferencias. Colabora en proyectos de código abierto, se conecta con otros científicos de datos y adquiere experiencia en diversas industrias cuando asistió a la Conferencia IBM TechXchange.

Experiencia en el Dominio

Entendiendo que el conocimiento del dominio es crucial, Lila elige un nicho que se alinea con sus intereses. Investiga a fondo varios dominios, incluyendo comercio electrónico, salud, finanzas y varios otros campos a los que podría aplicar sus habilidades en ciencia de datos de manera efectiva. Desde su maestría en economía, eligió el comercio electrónico como su dominio principal para conseguir una carrera en ciencia de datos.

Conseguir el Primer Empleo

Después de meses de preparación, Lila comenzó a postularse para puestos de científica de datos. Ella adapta su currículum para resaltar sus habilidades y

proyectos relevantes. Su portafolio en línea muestra sus capacidades y demuestra su compromiso con el campo.

Enfoque de Lila para Trabajar en Su Primera Tarea como Científica de Datos

Como científica de datos junior recién contratada en una empresa de retail, Lila utiliza los conocimientos derivados de los datos para mejorar el servicio al cliente. Su primera tarea consiste en analizar los datos de los clientes para identificar patrones y anomalías que podrían afectar el servicio al cliente. Ella utiliza el análisis de datos para mejorar la experiencia general del cliente.

Selección y Obtención de Conjuntos de Datos

En la fase inicial de su viaje en ciencia de datos, Lila enfrentó el desafío de seleccionar un conjunto de datos adecuado y obtenerlo de diferentes fuentes. Aparte de los datos históricos disponibles para las organizaciones durante los últimos cuatro años, exploró varios repositorios, sitios web y bases de datos para encontrar los conjuntos de datos correctos para su proyecto. Al recopilar datos de diversas fuentes, Lila se encontró con otro punto de decisión crucial. Tenía que decidir cómo armonizar e integrar estos conjuntos de datos dispares en un todo cohesivo. Se puso en contacto con profesionales de productos, ingenieros de datos y especialistas en el dominio, buscando su opinión y experiencia en la fusión de conjuntos de datos.

Comprensión y Limpieza de Datos

Lila comienza importando el conjunto de datos en su entorno de análisis de datos utilizando Python y SQL. Carga los datos y examina las primeras filas para entender su estructura y contenido. Al adquirir el conjunto de datos, Lila se enfrenta a su primer desafío: la limpieza de datos. Lila verifica si hay valores faltantes, duplicados y atípicos en el conjunto de datos. Aborda los datos faltantes imputando o eliminando filas o columnas con valores faltantes. Los valores atípicos se identifican y tratan adecuadamente según su impacto en el análisis.

Análisis Exploratorio de Datos (AED)

A medida que se adentra en el análisis exploratorio de datos, Lila enfrenta numerosas elecciones. Debe determinar qué estadísticas resumen, visualizaciones y análisis de distribución revelarán mejor información sobre el comportamiento del cliente y las tendencias de ventas. Cada elección que hace durante el AED influye en la historia que los datos contarán. Lila realiza AED para obtener información sobre el conjunto

de datos. Genera estadísticas resumen y visualizaciones (histogramas, gráficos de dispersión) y explora la distribución de variables. El AED le ayuda a comprender el comportamiento del cliente, los productos populares y las tendencias de ventas.

Ingeniería de características

Lila reconoce el potencial de la ingeniería de características para mejorar su análisis. Evalúa si la creación de nuevas características, como el cálculo de los montos totales de compra, mejorará la utilidad del conjunto de datos para su proyecto.

Análisis Estadístico, Aprendizaje Automático

Lila evalúa si son necesarios tests estadísticos o algoritmos de aprendizaje automático. Ella emplea análisis de regresión para entender las relaciones entre variables y explora modelos de aprendizaje automático para tareas de pronóstico de demanda o segmentación de clientes. Lila también realiza pruebas estadísticas para descubrir patrones en los datos. Ella utiliza análisis de regresión para entender el impacto del precio unitario en las ventas.

Presentación e Informes

Al culminar su análisis, Lila enfrenta el desafío de presentar sus hallazgos. Lila compila su análisis y hallazgos utilizando un Jupyter Notebook en un informe y presentación completos. Destaca ideas y recomendaciones prácticas para los interesados de la plataforma de comercio electrónico.

Aprendizaje Continuo

Después de completar su primer proyecto, Lila continúa perfeccionando sus habilidades, explora conjuntos de datos más complejos y enfrenta tareas de ciencia de datos cada vez más desafiantes.

Habilidades en Aprendizaje Automático

Aunque Lila tomó un curso introductorio sobre Aprendizaje Automático como parte del Certificado Profesional en Ciencia de Datos de IBM, el campo le intriga y quiere desarrollar sus habilidades aún más tomando el Certificado Profesional en Aprendizaje Automático de IBM. Identificó conjuntos de datos del Repositorio de Aprendizaje Automático en el curso y experimentó con varios algoritmos. Lila se sumerge en el aprendizaje automático para sobresalir como científica de datos, donde estudia varios algoritmos, como la regresión lineal, los árboles de decisión y

los modelos de aprendizaje profundo. Continúa adquiriendo experiencia en la selección y ajuste de algoritmos en función de problemas de datos específicos.