

Vicerrectoría de Educación Digital
Dirección de Innovación Digital

Guía para el profesor

Ingeniería de datos masivos



Índice

Certificados	3
¿Certificado o certificación?	3
Certificado en Tecnologías de Big Data	4
Competencia global del certificado	4
Competencia del curso	4
Metodología del curso.....	4
Temario	6
Bibliografía y recursos especiales	6
Evaluación y agendas.....	7
Notas de enseñanza	9

Certificados

Para entender la importancia del curso en el que usted será **facilitador**, es necesario ofrecer un mayor contexto sobre el programa de certificados de la Universidad Tecmilenio, pues son parte medular del modelo educativo basado en **aprender haciendo** y en **brindar una experiencia educativa a la medida de los alumnos**.

Un certificado es un programa académico corto compuesto de varias materias, incorporado en la segunda mitad del plan de estudios de profesional, que busca desarrollar competencias muy específicas en el alumno y lo prepara para desempeñarse de la mejor manera para un empleo.

saber + hacer + bien

Con este enfoque, buscamos que los egresados de profesional además de saber (tener un conocimiento teórico), también sean capaces de hacer (tener la habilidad de realizar una tarea) y de saber-hacer (entender lo que se hace y tener la capacidad para hacerlo de la mejor forma).

En Universidad Tecmilenio, **aprender haciendo** significa que el participante cursará certificados en los que desarrollará competencias disciplinares de especialidad que son valoradas por el mercado laboral, convirtiéndose en un profesional altamente competente y elevando así su índice de empleabilidad.



La mayoría de nuestros certificados se componen de cuatro materias en promedio, las cuales llevan un seguimiento lógico y finalizan con un proyecto de largo alcance con un alto nivel de complejidad (última materia). Asimismo, la realización correcta del proyecto integrador demostrará el dominio de la competencia global declarada en cada certificado.

¿Certificado o certificación?

Es muy importante tener en claro que un certificado y una certificación son dos cosas distintas. Un certificado es un reconocimiento formal que otorga internamente la Universidad Tecmilenio a los estudiantes que demuestren la aprobación de las materias correspondientes y el haber adquirido la competencia global del certificado.

Por su parte, la certificación también es un reconocimiento; sin embargo, esta se obtiene a través de la acreditación de un curso específico del programa académico de la Universidad y aprobando un examen de suficiencia aplicado por una entidad acreditadora externa (mapas mentales, idiomas, uso de software, etcétera).

Su trabajo como docente facilitador es muy importante para nosotros, por esa razón se le agradece su conocimiento y experiencia en la impartición de este certificado. A continuación, podrá revisar información detallada del curso que impartirá.

Certificado en Tecnologías de Big Data

El certificado en Tecnologías de Big Data se compone de tres cursos y un proyecto integrador con la siguiente progresión:



Como se puede apreciar, este curso de Fundamentos de programación para Big Data es el primer curso del certificado en Tecnologías de Big Data. Por lo mismo, es importante que como facilitador verifique que sus estudiantes hayan aprobado los cursos anteriores, pues de no haberlo hecho se podrá ver afectado el aprovechamiento académico del curso.

Competencia global del certificado

Al finalizar el certificado en Tecnologías de Big Data el participante deberá haber desarrollado y adquirido la siguiente competencia global en toda su extensión:

Diseña e implementa una infraestructura de Big Data y analíticos eficiente y confiable, acorde a las necesidades de cualquier tipo de organización multinacional.

Competencia del curso

La competencia específica que el participante habrá de obtener al aprobar satisfactoriamente el curso de **Ingeniería de datos masivos** es la siguiente en toda su extensión:

Diseña interfaces de visualización de la información a través de distintos sistemas para una toma de decisiones basada en los datos recolectados.

Metodología del curso

En este curso de Fundamentos de programación para Big Data se revisarán 10 temas.

En cada tema el participante encontrará:

- Una breve explicación del tema en PDF que ayudará al estudiante a ampliar su conocimiento.
- Una serie de lecturas, videos, podcasts y MOOC recomendados para una mejor comprensión de los temas.

GUÍA PARA EL PROFESOR

- Un video al inicio de cada tema para que el alumno pueda conocer las diferentes aplicaciones de Big Data.
- Una tarea o actividad de aprendizaje (no evaluable, pero si requisitoria), cuyo propósito es aplicar y experimentar los conceptos estudiados.

A lo largo del curso el participante debe trabajar en lo siguiente:

- Diez actividades
- Dos evidencias.

Actividades

Las actividades deben enviarse a través de la plataforma Canvas en la fecha indicada.

Evidencias

Las dos evidencias de este curso consisten en desarrollar un proyecto de ciencia de datos para la empresa *Instacart*. La primera evidencia incluye la entrega de las etapas correspondientes al entendimiento del negocio y de los datos. La segunda evidencia incluye la entrega de las etapas de preparación y modelación de los datos. Es importante revisar la agenda del curso, pues la mayoría de las **evidencias requieren entregas de avances** que los alumnos deben realizar conforme avanza el periodo académico.

NOTA

Es de suma importancia que enfatice en los participantes guardar todos los trabajos y productos que generen durante el curso (actividades, tareas, evidencias, etc.). Esto les servirá para conformar un portafolio personal de proyectos, así como para la elaboración de su proyecto integrador (último curso del certificado). Para ello, se le solicita colocar un aviso en Canvas (sección anuncios), tomando como referencia el siguiente texto:

Estimado participante, recuerda guardar siempre una copia digital de todos los trabajos, actividades y evidencias que realices en tus cursos. Contar con estos documentos te será de utilidad especialmente para dos fines:

1. Conformar un portafolio personal de proyectos, que te servirá como un medio importante para enriquecer tu proyección profesional.
2. Poder elaborar el proyecto integrador de tu certificado (última materia).

Por lo tanto, asegúrate de respaldar todos tus documentos localmente en un disco duro (computadora + USB flash drive), y de preferencia, también almacenarlos en la nube (servicios como Dropbox y Google Drive).

Temario

Los temas que se abordarán en este curso de certificado son los siguientes:

- Tema 1.** Fundamentos de Data Science.
- Tema 2.** Data Mining para Big Data.
- Tema 3.** Data Modeling orientado a Big Data.
- Tema 4.** Categorías de grupos de datos para Big Data.
- Tema 5.** Tratamiento e información de los datos.
- Tema 6.** Infraestructura para la distribución y procesamiento de datos.
- Tema 7.** Métodos de agrupación de datos.
- Tema 8.** Técnicas de visualización de datos y sumas numéricas.
- Tema 9.** Correlaciones en Big Data.
- Tema 10.** Análisis de series temporales para Big Data.

Bibliografía y recursos especiales

Los libros de texto son los siguientes:

Grus, J. (2019). *Data Science from Scratch: First Principles with Python* (2nd Edition). [Kindle]. Estados Unidos. O'REILLY

ISBN-10: 1492041130

ISBN-13: 9781492041139

Burkov, A. (2019). *The Hundred-Page Machine Learning Book* [Kindle]. Andriy Burkov

ISBN-13: 978-1999579500

ISBN-10: 199957950X

Evaluación y agendas

La evaluación del curso se estructura de la siguiente manera:

Unidades	Instrumento Evaluador	Puntaje
2	Evidencia	40
4	Avances de evidencias	60
Total		100 puntos

Una vez que se hayan **validado fechas y valores con la información disponible en Servicios en Línea**, las evidencias se entregarán de acuerdo con la siguiente agenda:

Agenda	
Evidencia	Ponderación
Avance 1	15
Avance 2	15
Evidencia 1	20
Avance 1	15
Avance 2	15
Evidencia 2	20
Total	100

Banner

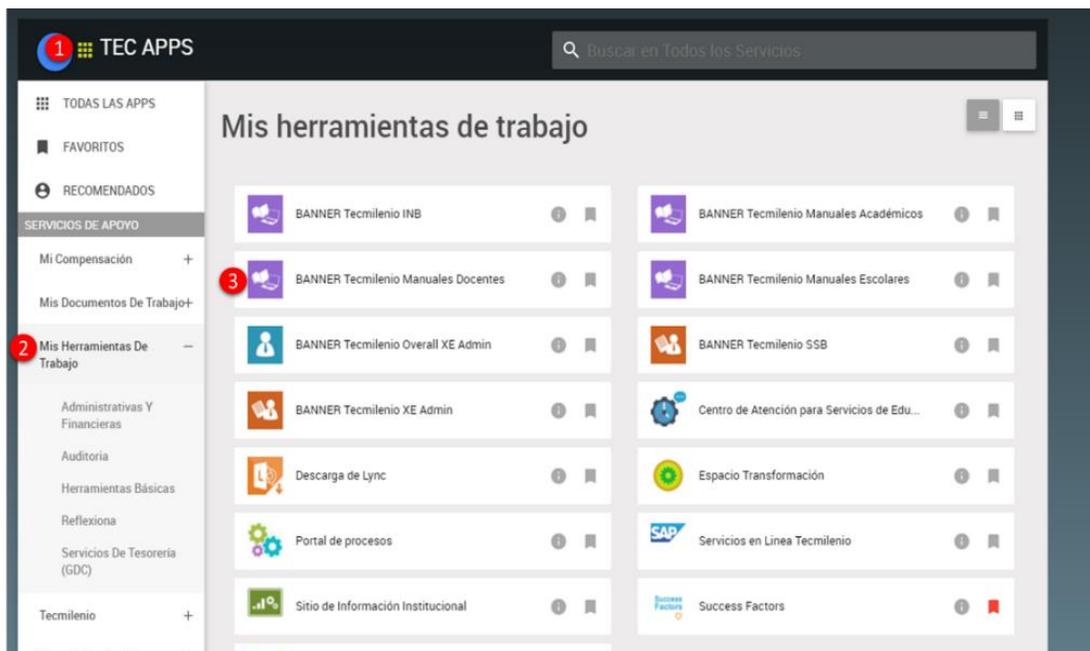
Estimado profesor, no olvide capturar las calificaciones de su grupo en las fechas indicadas.

Puede ver el manual para capturar calificaciones siguiendo esta ruta en MiEspacio:

- MiEspacio → TEC Apps → Mis Herramientas de Trabajo → BANNER Tecmilenio Manuales Docentes

Si imparte clase de manera presencial, puedes ver un manual para capturar inasistencias siguiendo esta ruta en MiEspacio:

- MiEspacio → TEC Apps → Mis Herramientas de Trabajo → BANNER Tecmilenio Manuales



Notas de enseñanza

Antes de impartir el curso, por favor revise de manera general los datos y conceptos proporcionados en el mismo, esto con el fin de detectar, y en su caso, poder actualizar o enriquecer previamente la información específica al tiempo en que se está impartiendo el curso.

Un aspecto de gran importancia en el desarrollo de los temas es su involucramiento como facilitador para propiciar que la competencia del curso se cumpla. Además, debe preparar a los participantes para que vayan desarrollando propuestas de soluciones innovadoras a problemas actuales propios del área de estudio.

Enseguida, puede revisar las notas de enseñanza generales y por tema para este curso.

Generalidades

Para la impartición de este curso se sugiere:

1. Revisar con tiempo la lista de entregables y la agenda en BANNER para saber en qué temas y semanas se deben realizar las evidencias y actividades.
2. Revisar periódicamente el foro de discusión en Canvas para resolver las preguntas e inquietudes de los participantes acerca de las actividades y la evidencia.
3. Motivar al alumno a participar y realizar sus actividades a tiempo.
4. Proveer retroalimentación constante de las actividades que realizan los participantes.
5. Elaborar una agenda y subirla a la plataforma para que los participantes puedan visualizar de manera esquemática los temas y actividades que deberán revisar cada semana.
6. Recordar (opcionalmente) a los participantes acerca de las entregas de sus actividades por medio de la sección de avisos de Canvas.
7. Enriquecer el curso con videos o lecturas adicionales.

Tema 1

Notas de enseñanza para el tema:

1. Analizar con los participantes la importancia de big data y la ciencia de datos en el caso de estudio de Amazon ubicado en la introducción del tema.
2. Invitar al participante a profundizar el caso de Amazon a través del libro de Bernard Marr (descarga gratuita), además de leer los otros seis casos que presenta la obra.
3. Reflexionar sobre los conceptos de big data y data science y sus campos de aplicación en las empresas en la actualidad y en el futuro.
4. Profundizar las cinco V de big data y el diagrama de Venn de Drew Conway.
5. Reflexionar con los participantes los diversos usos de las tecnologías de la inteligencia artificial, el aprendizaje automático, el aprendizaje profundo en el desarrollo de proyectos de data science, así como la minería e ingeniería de datos.
6. Analizar el origen y los ejemplos de los conceptos de la analítica de los negocios, científico de datos e inteligencia de negocios.

Tema 1

7. Invitar al participante a que revise los recursos de apoyo en su totalidad y que se involucre en investigar información adicional a la presentada.

Notas de enseñanza para la actividad:

1. Asegurar que el participante relacione correctamente los conceptos con las herramientas investigadas en la infografía.
2. Revisar que el participante incluya el proveedores y herramientas para cada tipo de tecnología señalada en la actividad.
3. Recomendar al participante que analice todas las opciones disponibles para generar la presentación electrónica y no solo en aquella en la que ha trabajado anteriormente.

Tema 2

Notas de enseñanza para el tema:

1. Compartir con los participantes las áreas donde se puede aplicar un proyecto de minería de datos.
2. Analizar con los participantes la definición y objetivo de utilizar la minería de datos.
3. Reflexionar las ventajas y desventajas de la minería de datos.
4. Profundizar en el ciclo de vida de Data Science y explicar su relación con la minería de datos (Data Mining).
5. Reflexionar sobre el proceso que existe en un proyecto de minería de datos.
6. Analizar a detalle la metodología CRISP DM y examinar algunos ejemplos de su aplicación.
7. Invitar al participante a que revise los recursos de apoyo en su totalidad y se involucre en investigar información adicional a la presentada.

Notas de enseñanza para la actividad:

1. Revisar que el participante incluya los criterios de evaluación solicitados para cada una de las herramientas que ha investigado.
2. Asegurar que el participante incluya fuentes confiables durante su investigación.
3. Validar que el participante relacione correctamente los conceptos del tema con las herramientas investigadas en la infografía.
4. Recomendar al participante analizar todas las opciones disponibles para generar la presentación electrónica y no solo en aquella en la que ha trabajado anteriormente.

Tema 3

Notas de enseñanza para el tema:

1. Reflexionar sobre los diversos esquemas existentes en la modelación de datos (adicional al modelo estrella), señalando las ventajas y desventajas de cada uno.
2. Profundizar sobre las ventajas, desventajas y características de modelo estrella, además de las tablas de hechos y dimensiones.
3. Analizar con los participantes la definición e importancia de la modelación de datos en Big Data.
4. Compartir ejemplos a los participantes sobre cada uno de los modelos de análisis de datos.
5. Invitar al participante a que pueda profundizar sobre cada uno de los modelos de análisis datos presentados.
6. Promover que el participante consulte los recursos adicionales propuestos.

Notas de enseñanza para la actividad:

1. Asegurar que el participante incluya respuestas claras y originales para cada una de las dos preguntas planteadas.
2. Revisar que el participante haga la dos entregas: el reporte y los Jupyter notebooks de cada uno de los modelos investigados.
3. Orientar a los participantes sobre el uso de las dos plataformas (Github y Kaggle) en la industria tecnológica hoy en día y las ventajas de realizar la investigación de Jupyter Notebooks.
4. Es el primer ejercicio del curso donde el participante volverá a utilizar Python, por lo que se debe recordar al participante tenga instalado lo necesario en su equipo de cómputo personal para poder abrir los Jupyter Notebooks.

Tema 4

Notas de enseñanza para el tema:

1. Analizar con los participantes la importancia de los datos como un nuevo activo en las organizaciones, además de las ventajas y desventajas que esto con lleva en una economía digital y globalizada.
2. Promover la curiosidad de los participantes para que lean y discutan en clase el artículo de David Peris “The world’s most valuable resource is no longer oil, but data”, para posteriormente compararlo contra un artículo en Internet de fuente confiable con un punto de vista opuesto.
3. Reflexionar sobre las infraestructuras necesarias existentes en el mercado para poder almacenar los datos generados en las organizaciones.
4. Compartir ejemplos sobre los datos en sus diferentes tipos de estructura, además de promover una comparación entre las tres mencionadas en el contenido del tema.
5. Analizar las consecuencias para las organizaciones que implica que el 80% de los datos sean no estructurados.
6. Invitar al participante a investigar la aplicación en la industria tecnológica de los tipos de datos de acuerdo a la perspectiva estadística.
7. Reflexionar sobre la diferencia entre los datos, la información y el conocimiento, esto le servirá cuando tengan que realizar la actividad del tema.

Notas de enseñanza para la actividad:

1. Orientar a que los participantes consulten el tutorial anexo en la actividad para que logren crearlo y subirlo exitosamente el video a YouTube. Los videos deben ser configurados para que sean de visualización privada, de esta forma evitaremos que otros alumnos puedan visualizarlo.
2. Asegurar que el participante incluya fuentes confiables durante su investigación.
3. Revisar que los participantes incluyan y relacione correctamente todos los elementos investigados en la videograbación, los cuales deben ser explicados de forma clara y precisa.
4. Validar que el participante realice la entrega del documento con la liga del video realizado y las referencias bibliográficas en formato APA.

Notas de enseñanza para el tema:

1. Realizar junto con los participantes diversos ejemplos sobre preparación de datos, las cuales serán de refuerzo previo al desarrollo de la actividad correspondiente al tema.
2. Analizar la mayor cantidad de herramientas para la realización del tratamiento de los datos, además de que los alumnos investiguen por su cuenta las más populares hoy en día.
3. Promover entre los participantes su curiosidad para que investiguen con más detalle los pasos presentes en el tratamiento de los datos mediante la óptica de al menos dos autores diferentes.
4. Analizar la ubicación de la preparación de los datos dentro de la metodología CRISP-DM, además de la etapa previa y posterior a este proceso.
5. Reflexionar junto con los participantes sobre la importancia de la preparación de datos y el tiempo que conlleva actualmente este proceso. Consulte el siguiente material:

Tripathi, G. (2017). *Why data preparation is an important part of data science?*. Recuperado de <https://medium.com/@gauravkth/why-data-preparation-is-an-important-part-of-data-science-e40e054f18eb>

SAS (2017). *Why Analytical Data Preparation is so Important*. Recuperado de <https://blogs.sas.com/content/hiddeninsights/2017/11/30/analytical-data-preparation-important/>

Brown, E. (2016). *Data Analytics – The importance of Data Preparation*. Recuperado de <https://ericbrown.com/data-analytics-data-preparation.htm>

6. Promover que el participante consulte los recursos adicionales propuestos.

Notas de enseñanza para la actividad:

1. Orientar a que los participantes consulten el video sobre preparación de datos señalado hacia el final del tema, ya que les será de utilidad para lograr una mayor comprensión de la actividad.
2. Asegurar que el participante pueda correr exitosamente ambos archivos anejados sobre preparación de datos en su computadora.
3. Solicitar que los practicantes incluyan comentarios claros durante el desarrollo del código en Python, con la finalidad de poder validar que logren entender cada uno de los pasos que están desarrollando.

Notas de enseñanza para el tema:

1. Analizar junto a los participantes la importancia de la estructura de datos y brindar consejos para que esta sea compatible con las necesidades actuales y futuras en la organización. Consulta el siguiente material:
Analytics8 (s.f.) *7 Tips for a Strong Data Infrastructure*. Recuperado de <https://www.analytics8.com/insights/7-tips-for-strong-data-infrastructure/>
2. Reflexionar sobre las tendencias para las próximas generaciones en la infraestructura de datos. Consulta el siguiente material:
Bullinger, P (s.f.). *6 Trends Shaping the Next Generation of Data Infrastructure*. Recuperado de <https://blog.westerndigital.com/trends-next-generation-data-infrastructure/>
3. Promover la curiosidad de los participantes para que investigue a detalle los diferentes tipos de almacenamientos de datos, lo cual servirá de complemento a la actividad del tema.
4. Invitar al participante a participar en clase discutiendo y analizando a Hadoop y Spark como sistemas distribuidos, en qué consisten, su importancia y utilización. Consulte el siguiente material
AWS (s.f.) *Apache Spark*. Recuperado de <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-spark.html>
SAS (s.f.). *Hadoop ¿Qué es y por qué es importante?* Recuperado de https://www.sas.com/es_mx/insights/big-data/hadoop.html
5. Reflexionar sobre las herramientas de analítica en la nube disponibles por los diferentes proveedores, compartiendo ejemplos. Consulte el siguiente material:
 - a. Langen, C. (2017). *Interacting with AWS S3 using Python in a Jupyter notebook*. Recuperado de <http://carolynlangen.com/2017/11/22/interacting-with-aws-s3-using-python-in-a-jupyter-notebook/>
 - b. Megler, V. (2018). *Access Amazon S3 data managed by AWS Glue Data Catalog from Amazon SageMaker notebooks*. Recuperado de <https://aws.amazon.com/blogs/machine-learning/access-amazon-s3-data-managed-by-aws-glue-data-catalog-from-amazon-sagemaker-notebooks/>
6. Promover que el participante consulte los recursos adicionales propuestos.

Notas de enseñanza para la actividad:

1. Asegurar que el participante haya creado una cuenta de AWS-educate días antes de entregar la actividad.
2. Orientar a que el participante relacione correctamente los productos, cursos y materiales de los conceptos señalados.
3. Revisar que el participante incluya la descripción y evidencias de cada concepto señalado en la actividad, además de explicar su relación.
4. Recomendar al participante analizar todas las opciones disponibles para generar la presentación electrónica y no solo en aquella en la que ha trabajado anteriormente.

Notas de enseñanza para el tema:

1. Promover entre los participantes su curiosidad para que investiguen los modelos de agrupación que existentes (además de los dos explicados en el tema), con el objetivo que puedan tener un panorama amplio sobre este tópico.
2. Compartir ejemplos donde se utiliza la agrupación de datos para diferentes problemáticas en diversas organizaciones (más allá de los presentados en el tema), por ejemplo, en logística, atención a clientes, servicios bancarios, etc.
3. Realizar junto con el participante diversos ejemplos de agrupación de datos en Python tomando como ejemplo el realizado dentro del video al final de tema. Además, será de utilidad cuando ellos realicen la actividad solicitada.
4. Analizar con los participantes el procedimiento a detalle para realizar la agrupación de datos mediante el algoritmo K-means, incluyendo el cálculo de las distancias y el uso del método Elbow.
5. Reflexionar sobre el modelo de agrupación jerárquica, los datos que se requieren para llevarla a cabo, su funcionamiento, medidas de distancia, criterios de vinculación, etc.

Notas de enseñanza para la actividad:

1. Orientar a que los participantes consulten el video sobre agrupación de datos señalado hacia el final del tema, ya que les será de utilidad para lograr una mayor comprensión de la actividad.
2. Asegurar que el participante pueda correr exitosamente ambos archivos anexados sobre agrupación de datos en su computadora.
3. Solicitar que los practicantes incluyan comentarios claros durante el desarrollo del código en Python, con la finalidad de poder validar que logren entender cada uno de los pasos que están desarrollando.

Notas de enseñanza para el tema:

1. Invitar al participante que pueda profundizar en conocer el uso de las herramientas de visualización como Power BI y Qlik. (ambas permiten obtener gratis versiones de prueba). Incluso Qlik mediante Qlik View permite en su página web poder visualizar y manipular diversas pruebas (demos).
2. Compartir ejemplos de la aplicación de la inteligencia de negocios (Business Intelligence) en distintos tipos de organizaciones, por ejemplo, parques de atracciones, consultoras de mercadotecnia, deportes, etc.
3. Analizar la importancia de la visualización de datos como una ventaja competitiva dentro de las organizaciones, sus alcances y las desventajas cuando no se utiliza adecuadamente o simplemente no se lleva a cabo.
4. Reflexionar sobre la relación entre los KPI y Big Data, tomando como consulta el siguiente artículo:

Marr, B. (.s.f) What Is The Relationship Between KPIs And Big Data? Recuperado de <https://www.bernardmarr.com/default.asp?contentID=1790>

5. Analizar los diferentes tipos de dashboards presentes en las diversas herramientas existentes para la visualización de datos, además de presentar (por ejemplo, en Power BI, Qlik, Google Data Studio, Bilbeo, etc), los cuales se pueden encontrar en la siguiente liga:

Predictive Analytics Today (s.f) 77 open source, free and top dashboards software. Recuperado de <https://www.predictiveanalyticstoday.com/open-source-dashboard-software/>

6. Realizar junto con los participantes ejemplos de cómo hacer sumas numéricas en Python y Tableau, además de utilizar Matplotlib de Python para generar gráficos.
7. Analizar con los participantes diferentes ejemplos de storytelling, dentro del siguiente portal podrán encontrar una gran cantidad (además de ejercicios y retos):
8. Storytelling with data(s.f.). Gallery. Recuperado de <http://www.storytellingwithdata.com/gallery>

Notas de enseñanza para la actividad:

1. Asegurar que el participante logre crear de forma exitosa la cuenta en Tableau Public y lo tenga instalado en su equipo de cómputo personal.
2. Solicitar a los participantes que deben realizar una lectura detallada de los datos proporcionados con la finalidad que puedan decidir una perspectiva de análisis visual de los datos, ya que de esto depende las visualizaciones, el dashboard y la historia correspondiente de la actividad.
3. Solicitar que los participantes consulten los videos sobre el uso de Tableau desarrollados dentro del tema, ya que ambos usan la misma base de datos que necesitarán para desarrollar la actividad.
4. Validar que la perspectiva de análisis visual de los datos, las visualizaciones, dashboard y la historia sean completamente diferentes a las presentadas en ambos videos dentro del tema.

Notas de enseñanza para el tema:

1. Analizar junto con los participantes el cómo Netflix utiliza el análisis de datos mediante correlaciones y patrones de consumo para seleccionar películas y crear contenido que le permite tomar decisiones que les generan grandes utilidades. Consulte el siguiente material:

Neilpatel (s.f.). *How Netflix Uses Analytics To Select Movies, Create Content, and Make Multimillion Dollar Decisions*. Recuperado de <https://neilpatel.com/blog/how-netflix-uses-analytics/>

2. Reflexionar sobre la importancia de la correlación dentro de Big Data, explicando la diferencia entre la causalidad y la correlación mediante ejemplos adicionales al presentado en el desarrollo del tema. Puede tomar los ejemplos presentados en el siguiente artículo:

Singh, S. (2018). *Why correlation does not imply causation?*. Recuperado de <https://towardsdatascience.com/why-correlation-does-not-imply-causation-5b99790df07e>

3. Promover entre los participantes su curiosidad para que investiguen las diferentes herramientas a utilizar para el análisis de correlaciones, con el objetivo de contar con un mejor panorama para que a futuro puedan explorarlas para resolver diferentes situaciones en su campo laboral.
4. Realizar junto con los participantes diversos ejemplos en Python sobre correlaciones, las cuales serán de refuerzo previo al desarrollo de la actividad correspondiente al tema.
5. Compartir ejemplos sobre los diversos tipos de correlaciones, tomando estas dos publicaciones:

Investopedia (2019). *What Does it Mean if the Correlation Coefficient is Positive, Negative, or Zero?* Recuperado de <https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp>

Ghosh, M. (2018). *Explaining Correlation To A Newbie To Data Analytics*. Recuperado de <https://analyticstraining.com/explaining-correlation-to-a-newbie-to-data-analytics/>

6. Promover que el participante consulte los recursos adicionales propuestos.

Notas de enseñanza para la actividad:

1. Asegurar que el participante pueda correr exitosamente ambos archivos anexados sobre correlación de datos en su computadora.
2. Orientar a que los participantes consulten el video sobre correlación de datos señalado hacia el final del tema, ya que les será de utilidad para lograr una mayor comprensión de la actividad.
3. Solicitar que los practicantes incluyan interpretaciones precisas y claras sobre las correlaciones realizadas.

Notas de enseñanza para el tema:

1. Analizar junto con los participantes el uso del análisis de las series de tiempo en las diferentes ramas como los negocios, la ciencia, finanzas, salud, entre otros. Consulte el siguiente material:

Merrett, R. (2019). *Time Series – the Quintillion Business Applications You Forgot About*. Recuperado de <https://blog.datasciencedojo.com/time-series-business-applications/>

Statswork. (2019). *Application of time series analysis in financial economics*. Recuperado de <https://medium.com/@statswork/application-of-time-series-analysis-in-financial-economics-c1a7e1b69d04>

2. Compartir ejemplos del uso de los pronósticos en las series de tiempo. Compartimos algunos ejemplos:

Tema 10

Contreras, A. & Atziry, C. & Martínez, J. & Sánchez, D. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. Recuperado de <https://www.sciencedirect.com/science/article/pii/S0123592316300754>

3. Montes, E. & Calvete, F. & Mantilla, C. (2016). Aplicación de series de tiempo en la realización de pronósticos de producción. Recuperado de <https://revistas.uis.edu.co/index.php/revistafuentes/article/view/5595/5776>
4. Realizar junto con los participantes diversos ejemplos sobre predicciones de series de tiempo con los modelos ARIMA en Python, las cuales serán de refuerzo previo al desarrollo de la actividad correspondiente al tema.
5. Realizar junto con los participantes diversos ejemplos sobre análisis de series de tiempo, las cuales serán de refuerzo previo al desarrollo de la actividad correspondiente al tema.
6. Reflexionar sobre la definición de las series de tiempo, así como los componentes presentes dentro del análisis de las series de tiempo.
7. Promover entre los participantes su curiosidad para que investiguen con más detalle el proceso completo para realizar un análisis de series de tiempo.
8. Promover que el participante consulte los recursos adicionales propuestos.

Notas de enseñanza para la actividad:

1. Asegurar que el participante pueda correr exitosamente ambos archivos anexados sobre correlación de datos en su computadora.
2. Orientar a que los participantes consulten los videos sobre correlación de datos señalados en el tema, ya que les será de utilidad para lograr una mayor comprensión de la actividad.
3. Solicitar que los participantes incluyan los comentarios y conclusiones sobre los análisis de series de tiempo realizados, incluyendo todo lo que se solicita.

Evidencia 1

Notas de enseñanza para la evidencia 1:

1. La primera evidencia del curso se divide en tres avances, los cuales deberán ser entregados en tres fechas distintas.
2. La finalidad de la entrega en la primera evidencia es que el participante desarrolle un proyecto de ciencia de datos abarcando las etapas de entendimiento del negocio y de los datos, finalizando con una reflexión sobre el proyecto y una presentación de los hallazgos.
3. Las actividades desarrolladas en el curso por el participante (temas 1 al 6) son requisito para que este pueda realizar la primera evidencia. Dentro de los criterios de evaluación de cada avance existe un apartado donde se otorgan puntos por la entrega de cada actividad señalada como requisito.
4. Enfatizar al participante la instalación de Anaconda para lograr trabajar con Jupyter Notebooks de Python, la cual le ayudará también para desarrollar otras actividades dentro del curso.

Evidencia 1

5. El primer avance corresponde hasta el sexto punto, donde el participante debe investigar sobre la empresa, descargar la base de datos señalada, además de elaborar un diccionario y un diagrama de datos para realizar un análisis descriptivo.
6. Lea bien las instrucciones que debe de seguir el participante, de encontrar algún error o recursos fuera de servicio o inexistentes, favor de comunicarlo inmediatamente para que se puedan hacer los ajustes necesarios y de preferencia proporcione los recursos que puedan ir en sustitución de lo erróneo o faltante.
7. Señalar al participante la importancia de incluir la descripción de los objetivos del proyecto y la descripción de los datos disponibles.
8. El segundo avance corresponde hasta el décimo punto, donde el participante llevará a cabo un análisis exploratorio para lograr localizar hallazgos en los datos y presentarlos mediante diversas visualizaciones, partiendo de las preguntas señaladas.
9. La entrega final de la primera evidencia cubre la presentación de los hallazgos además de una reflexión del proyecto, dando respuesta a cada una de las preguntas señaladas.
10. Señalar al participante si desea utilizar Tableau para realizar las visualizaciones, pueda revisar el tema 8 del curso o bien indicarle la sección de recursos.

Evidencia 2

Notas de enseñanza para la evidencia 2:

1. La segunda evidencia del curso se divide en tres avances, los cuales deberán ser entregados en tres fechas distintas.
2. Las actividades desarrolladas en el curso por el participante (temas 7 al 10) son requisito para que este pueda realizar la evidencia. Dentro de los criterios de evaluación, existe un apartado donde se otorgan puntos por la entrega de cada actividad señalada como requisito.
3. Las tres entregas de esta evidencia se relacionan con la evidencia anterior, debido a que el participante debe continuar trabajando en el desarrollo del proyecto de ciencia de datos, ahora abarcando las etapas de preparación y modelación de los datos.
4. Enfatizar al participante la importancia de la investigación de los conceptos señalados en el primer avance de la evidencia, ya que son la base para utilizar los Jupyter Notebooks para realizar la preparación de los datos mediante los modelos K-means y análisis de canasta de mercado.
5. Lea bien las instrucciones que debe de seguir el participante, de encontrar algún error o recursos fuera de servicio o inexistentes, favor de comunicarlo inmediatamente para que se puedan hacer los ajustes necesarios y de preferencia proporcione los recursos que puedan ir en sustitución de lo erróneo o faltante.
6. Enfatizar al participante la importancia de seleccionar uno de los segmentos de clientes para lograr identificar sus características comunes, llevando a cabo una descripción de ellas.
7. En la entrega final de la evidencia se debe reforzar en la presentación de la propuesta de mejora para la operación de la empresa señalada y como mediante las herramientas trabajadas se puede lograr.
8. Señalar la importancia de revisar la rúbrica para la evidencia final por parte del participante.