

Bloque 2.

Tema 3:
Probabilidad y
estadística para
data science

¿Por qué es necesaria la probabilidad y la estadística para data science?

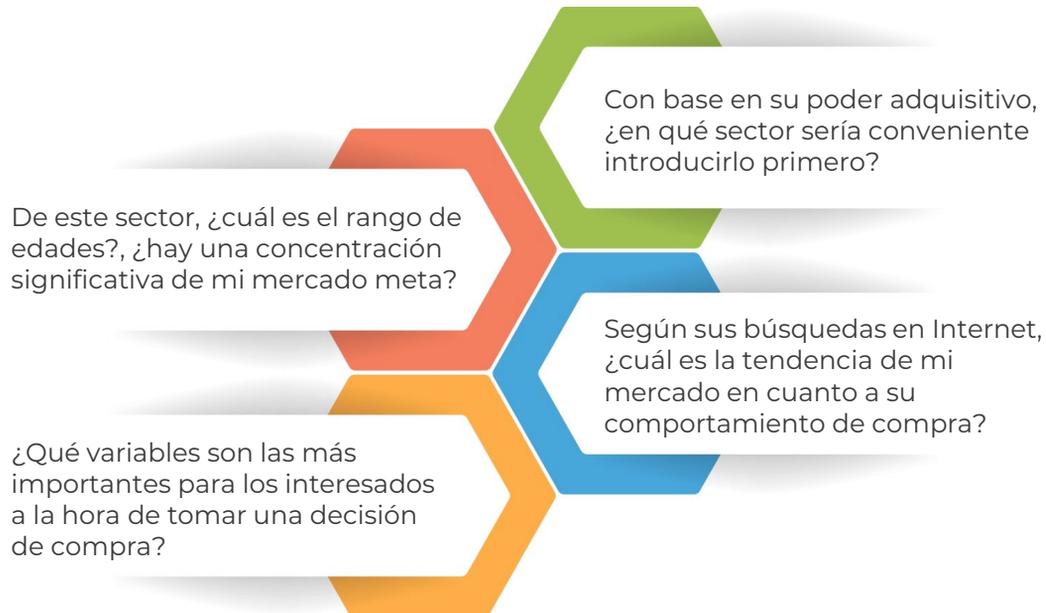


Data science busca la extracción de conocimiento para comprender fenómenos, sucesos o situaciones (Teate, 2015), y debido a esto, es necesario contar con un perfil que nos ayude a utilizar métodos matemáticos con el fin de comprender y describir dichos fenómenos.

Es aquí donde podemos apoyarnos en la probabilidad y en la estadística, ya que estas ciencias se basan en la recolección, análisis, interpretación, presentación y organización de los datos (Stanford University, s.f.).

Asimismo, la probabilidad y la estadística nos ayudan a resolver preguntas específicas, por ejemplo, digamos que buscamos vender un producto con características específicas para el sector femenino.

De acuerdo con la información recolectada y analizada de Internet a través de los métodos estadísticos y de probabilidad, se pueden establecer las siguientes preguntas:



Estadística

La estadística es la recopilación, análisis, interpretación, presentación y organización de los datos. Asimismo, es utilizada en un amplio campo de industrias y áreas, por lo tanto, no es sorprendente saber que los científicos de datos necesiten tener conocimientos de estadística.

Esto quiere decir que el análisis de datos requiere al menos de estadística descriptiva y de la teoría de la probabilidad, ya que estas teorías ayudan a tomar mejores decisiones comerciales basadas en los datos (Patel, 2018).

Asimismo, la estadística hace uso de las matemáticas para realizar análisis técnicos de datos, obteniendo así información concreta, ya que una visualización básica (como un gráfico de barras) puede brindarte información importante, sin embargo, con las estadísticas se pueden operar los datos de una manera más orientada a la información para conocer a detalle y profundidad sobre su estructura y su comportamiento.

Por lo tanto, existen cinco conceptos básicos de la estadística que los científicos de datos necesitan saber (Seif, 2018):

- 1 Características de la estadística.
- 2 Distribución de probabilidad.
- 3 Reducción de dimensionalidad.
- 4 Muestreo.
- 5 Estadística bayesiana.

Para saber más sobre **estadística**, te invitamos a ver el siguiente video:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.



GeoEngineerings Society. (2019, 19 de enero). *What is Statistics?* [Archivo de video]. Recuperado de <https://bit.ly/2StqxZj>

Para saber más sobre **cómo aplicar estadísticas en data science**, te invitamos a ver el siguiente video:



ACADGILD. (2018, 18 de mayo). *Statistics for Data Science 2018 Part 1 | Statistics Tutorial For Beginners | Data Science Tutorial* [Archivo de video]. Recuperado de <https://bit.ly/2NrBU4l>



Probabilidad

La probabilidad es la posibilidad de que algo suceda y se cuantifica como un número entre 0 y 1, donde 0 indica imposibilidad y 1 indica certeza.

Asimismo, tener conocimiento sobre los conceptos de probabilidad nos ayuda a formular juicios informados sobre lo que es probable que suceda con base en un patrón de datos previamente recopilados o de una estimación.

Data science utiliza inferencias estadísticas para predecir o analizar tendencias a partir de datos, empleando del mismo modo distribuciones de probabilidad. Por lo tanto, conocer sobre la probabilidad y sus aplicaciones son importantes para trabajar eficazmente los problemas de la ciencia de datos.

Para saber más sobre **probabilidad**, te invitamos a ver el siguiente video:



GeoEngineerings Society. (2018, 22 de diciembre). *What is Probability?* [Archivo de video]. Recuperado de <https://bit.ly/2Zq89SN>

Para saber más sobre **cómo aplicar la probabilidad en data science**, te invitamos a ver el siguiente video:



Rathi, A. (2018, 31 de julio). *Probability for Data Science* [Archivo de video]. Recuperado de <https://www.linkedin.com/pulse/probability-statistics-data-science-series-ankit-rathi>

Referencias

- Patel, A. (2018). *How to self-learn statistics of data science*. Recuperado de <https://medium.com/ml-research-lab/how-to-self-learn-statistics-of-data-science-c05db1f7cfc3>
- Seif, G. (2018). *The 5 Basic Statistics Concepts Data Scientists Need to Know*. Recuperado de <https://towardsdatascience.com/the-5-basic-statistics-concepts-data-scientists-need-to-know-2c96740377ae>
- Stanford University. (s.f.). *Statistics and Data Science*. Recuperado de <https://stanford.io/2N25iLg>
- Teate, R. (2015). *What is Data Science?* Recuperado de <https://bit.ly/34ij5eq>

Para expandir tu conocimiento, te recomendamos los siguientes recursos adicionales:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.

Podcasts

Para saber más sobre **probabilidad y estadística en data science**, te recomendamos escuchar los siguientes podcasts:

- Fodor, J. (23 de enero de 2017). *Episode 79: Basic Concepts in Statistics* [Audio podcast]. Recuperado de <https://fods12.podbean.com/e/episode-79-basic-concepts-in-statistics/>
- Packmohr, P. (6 de septiembre de 2018). *Probability* [Audio podcast]. Recuperado de <https://bit.ly/379vFWL>

MOOC

Para aprender más sobre **probabilidad y estadística en data science**, te recomendamos cursar los siguientes MOOC:

- Harvard University. (s.f.). *Data Science: Probability*. Recuperado de <https://online-learning.harvard.edu/course/data-science-probability>
- University of Amsterdam. (s.f.). *Basic Statistics*. Recuperado de <https://es.coursera.org/learn/basic-statistics>
- UC San Diego. (s.f.). *Probability and Statistics in Data Science using Python*. Recuperado de <https://www.classcentral.com/course/edx-probability-and-statistics-in-data-science-using-python-8213>

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.