# AMAT Cursos Disponibles de Ciencia de Datos 2022

## Nombre: Kit básico para Ciencia de Datos con R (Tidyverse)

Duración: 24 HRS

## Objetivo:

Brindar al participante los elementos teóricos y prácticos básicos alrededor de la Ciencia de Datos. Aprenderá a distinguir las diferentes soluciones a problemas que pueden resolverse con algoritmos de machine learning y aprenderá a usar el conjunto de librerías en *R* más novedoso, estructurado y ampliamente usado para la manipulación, transformación y visualización de datos: “TIDYVERSE”.

## Alcances:

Al finalizar este curso el participante será capaz de consumir, manipular y visualizar información proveniente de diversas fuentes de información para resolver problemas de propósito general asociados a los datos. Este curso brindará las bases para introducirse a cualquiera de los cursos de Ciencia de datos que se imparten en AMAT:

1. Data Science & Machine Learning (Aprendizaje supervisado I)
2. Data Science & Machine Learning (Aprendizaje no supervisado)

## Requisitos: Computadora con al menos 4Gb Ram

## Temario:

* Introducción a Ciencia de Datos (6 HRS)
  + ¿Qué es Ciencia de Datos?
  + Objetivo de la ciencia de datos
  + ¿Qué se requiere para hacer ciencia de datos?
  + Tipos de problemas que se pueden resolver
  + Tipos de algoritmos y aprendizaje
* Transformación y manipulación de estructuras de datos (4 HRS)
  + Lectura de archivos
  + Consultas de datos
  + Transformación de estructuras de datos
* Manipulación de texto (6 HRS)
  + Operaciones con cadenas
  + Expresiones regulares
  + Extracción de regexp
  + Filtros con regexp
  + Sustitución de regexp
* Manipulación de fechas y horas (2 HRS)
  + Creación de datos temporales
  + Transformación de datos temporales
  + Operaciones con datos temporales
* Análisis exploratorio y visualización interactiva de datos (6 HRS)
  + EDA y GEDA
  + Uso, decisión e implementación de técnicas gráficas
  + Visualización interactiva

## Nombre: Data Science & Machine Learning

## **(Aprendizaje Supervisado I**)

Duración: 32 HR

## Objetivo:

Brindar al participante un amplio conocimiento de los requisitos, alcances y toma de decisión alrededor de un proyecto de Ciencia de Datos. Aprenderá a través del conjunto de librerías más novedoso en *R* a crear, ajustar y afinar los mejores parámetros para la correcta implementación de diferentes tipos de modelos **PREDICTIVOS** de aprendizaje de máquina estadístico (Statistical Machine Learning). De entre los distintos modelos creados, sabrá identificar cuál modelo ofrece el mejor ajuste de acuerdo con el objetivo particular de cada problema a resolver.

## Alcances:

Al finalizar el módulo, el participante sabrá plantear un proyecto de ciencia de datos, desde sus requerimientos hasta sus alcances. Sabrá crear flujos de trabajo limpios y ordenados para crear poderosos modelos de Machine Learning. Este curso brindará las bases para introducirse al módulo intermedio de Ciencia de datos que se imparte en AMAT:

1. Data Science & Machine Learning (Aprendizaje Supervisado II)

## Requisitos:

## Computadora con al menos 4Gb Ram.

Instalación de R con al menos versión 4.1.0

Instalación de Rstudio con al menos versión 1.4

Kit básico para Ciencia de Datos con R (Tidyverse) ó

## Dominio de las funciones de manipulación y visualización de datos con Tidyverse en *R*

## Temario:

* Introducción a Ciencia de Datos (8 HRS)
  + ¿Qué es Ciencia de Datos?
  + Objetivo de la ciencia de datos
  + ¿Qué se requiere para hacer ciencia de datos?
  + Tipos de problemas que se pueden resolver
  + Tipos de algoritmos y aprendizaje
  + Ciclo de vida de un proyecto de Ciencia de Datos
  + Taller de Scoping
* Machine Learning: conceptos básicos (4 HRS)
  + ML y algoritmos
  + Análisis supervisado vs no supervisado
  + Sesgo y varianza
  + Pre-procesamiento e ingeniería de datos
  + Partición de datos: test & train
* Machine Learning: Modelos de aprendizaje supervisado (20 HRS)
  + Regresión
  + Regresión logística
  + Regresión lasso
  + Regresión ridge
  + ElasticNet
  + KNN
  + Árbol de decisión
  + Bagging (básico)
  + Random Forest
  + Comparación de modelos

## Nombre: Data Science & Machine Learning **(Aprendizaje Supervisado II)**

Duración: 32 HRS

## Objetivo y alcances:

Brindar al participante el conocimiento y habilidades para implementar modelos complejos de Machine Learning a través de un flujo de trabajo limpio, ordenado y sistematizado a mediante las librerías en *R* más novedosas que han sido desarrolladas hasta el momento. Al finalizar este curso, el participante será capáz de combinar distintas clases de modelos para dar una solución compleja y precisa a problemas predictivos. Aprenderá a cuantificar los problemas éticos asociados al sesgo o inequidad producidos por modelos de machine learning, así como su interpretación en el mundo productivo. Finalmente, se estudiará la manera de desarrollar un diseño de experimento para implementarse en el ámbito empresarial de modo que el participante pueda tomar mejores decisiones para contribuir en su ambiente laboral.

## Requisitos:

## Computadora con al menos 4Gb Ram.

Instalación de R con al menos versión 4.1.0

Instalación de Rstudio con al menos versión 1.4

Data Science & Machine Learning (Aprendizaje Supervisado I)

## Temario:

* Machine Learning (10 HRS)
  + Regresión polinomial
  + Regresión CPA
  + Imputación avanzada
  + SVM
  + Boosting
* Flujos de trabajo y ensamblaje de modelos (8 HRS)
  + Pipelines
  + Workflowsets
  + Stacking

* Sesgo e inequitad de modelos (4 HRS)
  + Cuantificación de sesgo
  + Cuantificación de inequidad

* Interpretación de modelos (4 HRS)
  + Interpretación local de modelos
* Aplicación a negocios (6 HRS)
  + Diseño de experimentos en campañas de retención
  + Bagging (básico)
  + Random Forest
  + Comparación de modelos

## Nombre: Data Science & Machine Learning

## **(Aprendizaje No Supervisado)**

Duración: 24 HRS

## Objetivo:

Brindar al participante un amplio conocimiento de los requisitos, alcances y toma de decisión alrededor de un proyecto de Ciencia de Datos. Aprenderá a través del conjunto de librerías más novedoso en *R* a crear, ajustar y afinar los mejores parámetros para la correcta implementación de diferentes tipos de modelos de aprendizaje de máquina estadístico no supervisado (Not Supervised Statistical Machine Learning) para la creación de índices, segmentación y visualización de datos multivariados. De entre los distintos modelos creados, sabrá identificar cuál modelo ofrece el mejor ajuste de acuerdo con el objetivo particular de cada problema a resolver.

## Alcances:

Al finalizar el módulo, el participante sabrá plantear un proyecto de ciencias de datos, desde sus requerimientos hasta sus alcances. Sabrá crear flujos de trabajo limpios y ordenados para crear poderosos modelos no supervisados de Machine Learning.

## Requisitos:

## Computadora con al menos 4Gb Ram.

Instalación de R con al menos versión 4.1.0

Instalación de Rstudio con al menos versión 1.4

Curso: Kit básico para Ciencia de Datos con *R (Tidyverse)* ó

## Dominio de las funciones de manipulación y visualización de datos con Tidyverse en *R*

## Temario:

* Introducción a Ciencia de Datos (4 HRS)
  + ¿Qué es Ciencia de Datos?
  + Objetivo de la ciencia de datos
  + ¿Qué se requiere para hacer ciencia de datos?
  + Tipos de problemas que se pueden resolver
  + Tipos de algoritmos y aprendizaje
  + Ciclo de vida de un proyecto de Ciencia de Datos
* Análisis de componentes principales (4 HRS)
  + Interpretación geométrica
  + Cálculo de PCA
  + Creación de índices
* Cluster Jerárquico (4 HRS)
  + Liga simple, compleja y promedio
  + Dendogramas & Heatmap
* Cluster no Jerárquico (6 HRS)
  + K-means
  + K-medoides
  + Clara
  + DBSCAN
* Clusterización y visualización geoespacial (6 HRS)
  + Lectura de datos espaciales
  + Transformación de datos espaciales
  + Clustering espacial
  + Visualización espacial

Data Science for Business:

**(Para tomadores de decisiones)**

Duración: 24 HRS

## Objetivo:

Brindar al participante un amplio conocimiento de los requisitos, alcances y toma de decisión alrededor de un proyecto de Ciencia de Datos desde su planteamiento del problema y hasta su implementación en campañas preventivas. Se obtendrán los elementos básicos para poder dirigir a un equipo técnico que resuelve problemas a partir de datos y aprenderá a tomar decisiones al trabajar con proveedores o clientes que ofrezcan o requieran esta clase de servicios analísticos.

## Alcances:

Al finalizar este curso el participante será capaz de autogestionarse o dirigir a un equipo conformado por analistas técnicos en la creación de una solución por analítica predictiva a través de la creación de modelos de machine learning.

## Requisitos: Ninguno

Temario:

* Introducción a Ciencia de Datos (4 HRS)
  + ¿Qué es Ciencia de Datos?
  + Objetivo de la ciencia de datos
  + ¿Qué se requiere para hacer ciencia de datos?
  + Tipos de problemas que se pueden resolver
  + Tipos de algoritmos y aprendizaje
* Ciclo de vida (4 HRS)
  + Ciclo de vida de un proyecto de Ciencia de Datos
  + Taller de Scoping
* Análisis Exploratorio (4 HRS)
  + GEDA / EDA
  + Análisis y visualización
* Machine Learning (Aprendizaje Supervisado) (4 HRS)
  + ML y algoritmos
  + Sesgo y Varianza
  + Estimación de errores
  + Partición de datos
  + Métricas de desempeño
  + Algoritmos
* Diseño de experimentos (2 HRS)
  + Diseño de experimentos para ML
  + Estimación de utilidad
* Machine Learning (Aprendizaje NO Supervisado) (4 HRS)
  + Clustering
  + K-means
  + PAM
  + BBSCAN