



SECRETARÍA DE
INNOVACIÓN

CIENCIA DE DATOS



SECRETARÍA DE
INNOVACIÓN





Agenda

Sesión 18/18

Ciencia de Datos

- Análisis de fases de la ciencia de datos
 - Recopilación
 - Limpieza
 - Procesamiento de datos
 - Visualización
 - Comunicación

Review de Código de Python.

Recordemos

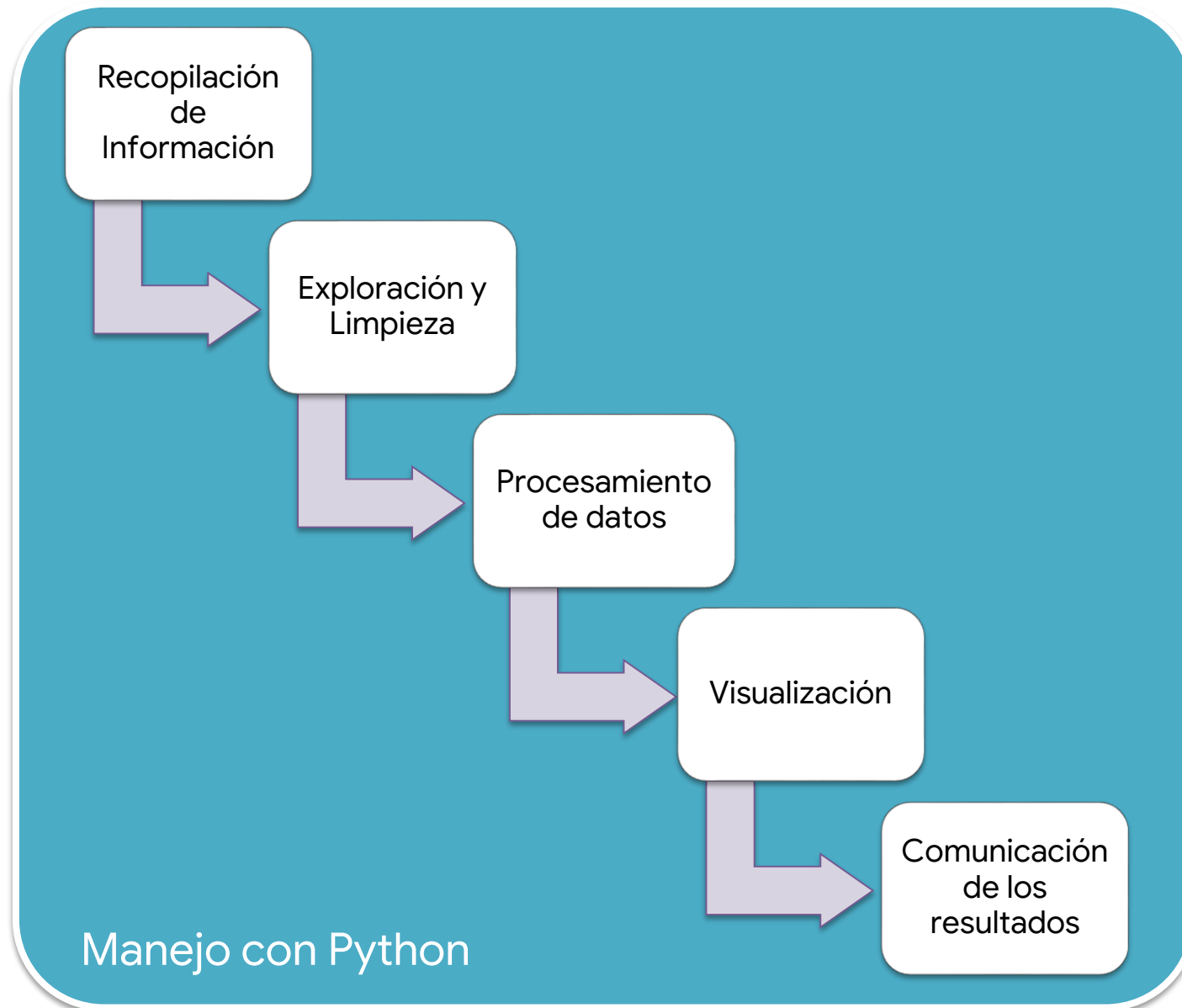
La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la **estadística**, la **minería de datos**, el **aprendizaje automático**, y la **analítica predictiva**.

Recordemos

La ciencia de datos busca realizar el proceso que permita obtener conocimiento de los datos que son estudiados.

Estos datos muchas veces poseen requerimientos ó necesidades que deben cumplirse para que la información sea integral. De esa manera se cumplirá que la información resultante al final del estudio tenga la veracidad y confiabilidad.

Recordemos sus fases



Recopilación de la Información

Es el proceso de obtener y medir información sobre variables específicas en un sistema establecido, que luego permite responder preguntas relevantes y evaluar resultados.

La recopilación de datos es un componente de la investigación en todos los campos de estudio, incluidas las ciencias físicas y sociales, las humanidades y los negocios, etc.

Recopilación de la Información

Herramientas de Recopilación de la información

- Formularios Web
- Investigaciones en Campo
- Big Data en la Nube
- Encuestas
- Bases de Datos
- Documentos de texto, csv, excel, etc.

Limpieza de datos

Es el primer paso mas critico en cualquier proyecto de inteligencia artificial y el Machine Learning. Es de los procesos mas importantes en el momento de hacer análisis de datos.

Es el proceso de detectar, corregir, eliminar registros corruptos o inexactos de un conjunto de registros, tablas o bases de datos.

Se refiere también a la identificación de partes incompletas, incorrectas, inexactas o irrelevantes de los datos para su posterior sustitución, modificación, eliminación de los datos sucios o poco precisos

Limpieza de datos

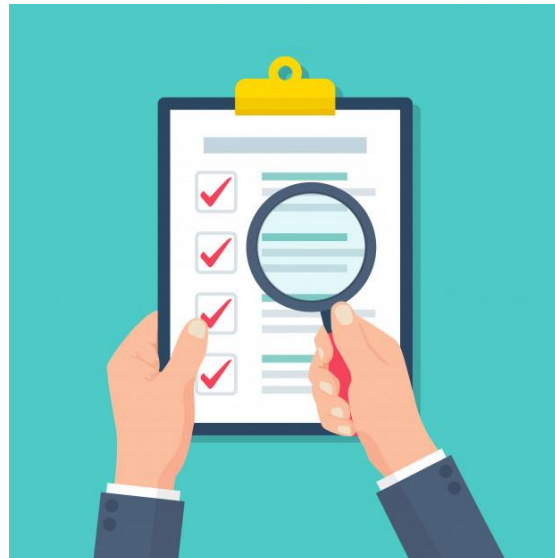
La calidad de datos debe cumplir con los siguientes requisitos:

- Exactitud
- Integridad
- Validez
- Consistencia
- Uniformidad
- Unicos

Limpieza de datos

Auditoría de Datos

Los datos son revisados con el empleo de métodos estadísticos de descubrir anomalías y contradicciones. Esto tarde o temprano da una indicación de las características de las anomalías y sus posiciones.



Procesamiento de los datos

El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

Además de la etapa de análisis en bruto, supone aspectos de gestión de datos y de bases de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de intereses, de consideraciones de la teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

Procesamiento de los datos

Proceso y Protocolos de la Minería de Datos

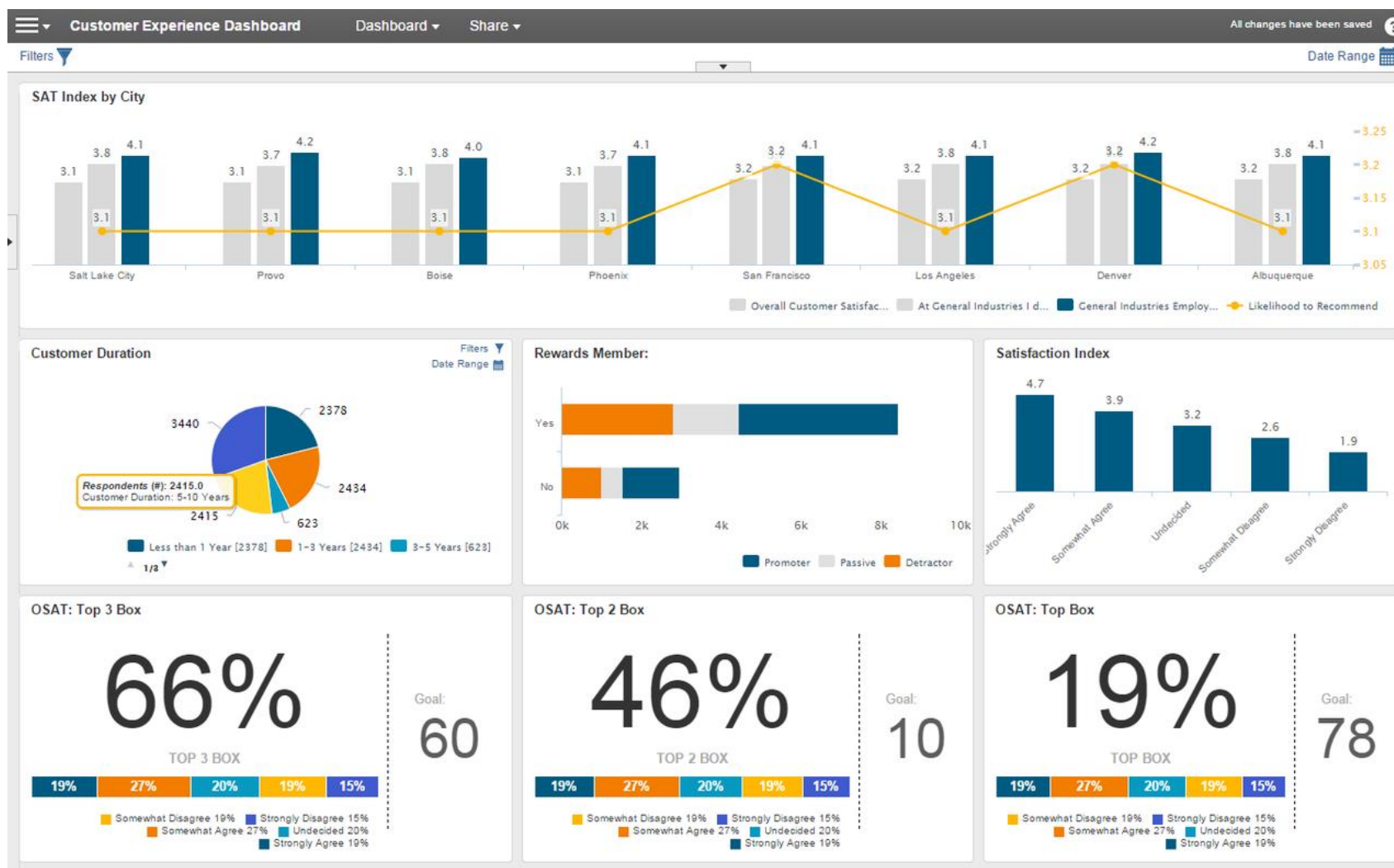
Un proceso, entendiéndose como conjunto de etapas sucesivas, típico de minería de datos consta de los siguientes pasos generales:

1. Selección del conjunto de datos
2. Análisis de las propiedades de los datos
3. Transformación del conjunto de datos de entrada
4. Selección y aplicación de la técnica de minería de datos
5. Extracción de conocimiento

Visualización de los datos

La visualización de datos es la representación gráfica de información y datos. Al utilizar elementos visuales como cuadros, gráficos y mapas, las herramientas de visualización de datos proporcionan una manera accesible de ver y comprender tendencias, valores atípicos y patrones en los datos.

Visualización de los datos



Visualización de los datos

La visualización de datos es otra forma de arte visual que capta nuestro interés y mantiene nuestros ojos en el mensaje. Cuando vemos un gráfico, vemos rápidamente las tendencias y los valores atípicos. Si podemos ver algo, lo interiorizamos rápidamente.

Es contar historias con un propósito. Si alguna vez haz visto una gigantesca hoja de cálculo de datos y no te fue posible ver una tendencia, sabes cuán eficaz puede ser una visualización.

Comunicación de los Resultados



RESUMEN DE CLASE



SECRETARÍA DE
INNOVACIÓN