## Introducción ANDREA

## Selección y Justificación de la base de datos.LISTO

## Búsqueda de la base de datos

## Los integrantes realizamos una búsqueda individual en diversas fuentes de bases de datos. Tras analizar las propuestas, a través de una sesión virtual, se acordó por unanimidad utilizar la base de datos que contiene los registros de alumnos en la Universidad de Chapingo.

## Justificación de la selección

La exploración al vuelo de los datos contenidos en la propuesta nos pareció completa, con los criterios necesarios y suficientes, lo que nos permitirá explorar y analizar las distintas hipótesis -que se describen más a detalle en los puntos 3 y 4 (Análisis exploratorio y Visualización de Datos). Con lo anterior, acordamos que: contiene un cúmulo lo suficientemente grande para ejercitar el manejo de grandes cantidades de registros (metadatos) y con las suficientes variables que nos brindaran bastas alternativas para indagar.

## 2.- Preparación y limpieza de datos. LISTO

## Resumen de los datos

Es una base de datos de la universidad de Chapingo, con 148431 datos, que recopila información del número de alumnos matriculados en los diferentes niveles que tienen disponibles en su programa académico; estos comprenden desde propedéutico hasta doctorado, en los años 2006 al 2021 en sus distintas unidades académicas.

Las columnas muestran información del año de ingreso, genero del alumno, a que grado ingresan, edad, nivel académico, categoría (becados, externos, media beca…) Unidad académica, sede, programa educativo, nacionalidad, estado de procedencia, lengua indígena y número de alumnos.

## Limpieza de datos

En primera instancia, detectamos la ausencia de ciertas especificaciones en la columna de *nacimiento*, por lo que se consideraron incompletos.

En el campo de *lengua indígena* solo en caso afirmativo había información, aunando al tipo de lengua se trataba. Aunado a esto, encontramos que existían registros que no eran individuales, sino de 2 o más con respecto al número de alumnos.

Por tanto, la corrección para esta situación fue, registrar uno solo con esas características. Se encontraron 264 registros en esa situación, que totalizan 556 alumnos, por lo que la limpieza da un total de 292 alumnos, que representan el 0.1963% del total de 148,722, por lo que esta limpieza no afecta la integridad ni es representativo del total.

Para el caso de las columnas de “Lengua indígena” y “fecha de nacimiento”, se rellenaron las celdas vacías con el dato “Null” y “N/A”, sin embargo, estas variables no formaron parte del análisis posterior.

Por último, en el campo de *nivel educativo,* fueron reclasificados los datos de las maestrías que estaban en nivel doctorado a “Maestrías”, así mismo el programa propedéutico al ser parte de la preparación para el “medio superior” se englobo en un solo nivel de medio superior.

## 3.- Análisis exploratorio de datos. JOSE CUESTAS (R)

## Análisis descriptivo.

(posibilidad de analizarlo con R o Phyton?)

## 4.- Visualización de datos. INCLUIR SI HAY MAS HIP

## Objetivos del estudio:

* General:

Identificar el nivel educativo, sede y programa educativo con mayor crecimiento en la matrícula de estudiantes, a través del análisis y visualización de los datos a lo largo del tiempo, para realizar una predicción de los próximos cinco años que nos permita estimar la demanda futura la cual permitirá apoyar la toma de decisiones estratégicas en materia de oferta académica.

* Particulares

En este análisis se quiere abordar diferentes puntos, por lo que parte de tres objetivos particulares:

1.- Conocer el año con el mayor número de alumnos matriculados, utilizando estadísticos descriptivos que permitan observar el efecto ocasionado por la pandemia del COVID-19 en 2020-2021.

## Hipótesis: CADA QUIEN INCLUIR

1.- Los años con menor alumnos matriculados fueron los años de la pandemia (2020-2021) en comparación a los años anteriores.

2.- El programa con mayor demanda es la preparatoria agrícola y los de menor demanda serán los de superior y doctorado, en los cuales se necesita una estrategia para promocionarlos.

## Plan de desarrollo: ANDREA.

Como se describió anteriormente, la etapa inicial consistió la limpieza exhaustiva de datos en Excel, con el objetivo de garantizar calidad y confiabilidad de la información, para ello se establecieron cuatro sencillos pero contundentes criterios:

1. Identificación y eliminación de datos nulos en variables significativas.
2. Corrección de valores atípicos fuera de un rango razonable.
3. Eliminación de valores duplicados en el alumnado.
4. Verificación de la consistencia en los datos, asegurando que cada registro pertenezca a la categoría.

Enseguida se realizó un análisis exploratorio de los datos analizados con R o Phyton? en el que se incluyeron….

Por último, se utiliza un visualizador de datos Tableau / Power BI, para poder comprender la información fácilmente y así llegar a una conclusión.

Se espera que este proceso de análisis permita identificar insertar objetivos

## Desarrollo:

Primero se hizo una limpieza de datos en Excel, en el cual todos los datos en blanco fueron remplazados por la leyenda “N/A”. Además de que todos los datos que no pertenecían a la categoría se reacomodaron a su categoría especifica (ej. Las maestrías que estaban en nivel doctorado fueron removidas a Maestría) y el programa propedéutico al ser parte de la preparación para el “medio superior” se englobo en un solo nivel.

Para visualizar el efecto pandemia se utiliza un gráfico lineal en donde se grafiquen el número de alumnos ingresados por año, en el que se puede filtrar por género.

Después se desarrolló un gráfico circular para mostrar cuales son los niveles en los que hay más ingresos, y posteriormente para completar esta información, se realiza un análisis más específico £de que programa educativo es el más concurrido por nivel.

## Resultados:

### Grafico lineal: Alumnos matriculados por año:

En este grafico se puede visualizar que al contrario de lo que se pensaba, el número de ingresos de los alumnos fue en incremento gradual desde el año 2006 hasta el año 2021, por lo que no se vio afectado por el inicio de la pandemia en el año 2020-2021, años en los que se registró la mayor cantidad de alumnos matriculados (10,807 y 11.199) respectivamente. Con un promedio de 9,277 alumnos inscritos.

Gráfico, Gráfico de líneas, Gráfico de dispersión

Descripción generada automáticamente

Con respecto al género, las mujeres ingresaron un promedio de 3,832:

Gráfico, Gráfico de líneas, Gráfico de dispersión

Descripción generada automáticamente

En comparación a los hombres estos fueron 5, 445:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

por lo tanto, hay un ingreso mayor en hombres que en mujeres.

### Grafico circular: Numero de ingresos por niveles:

En este segundo gráfico se puede observar el Numero de ingresos en nuestros distintos niveles académicos. Se observa que el mayor número de ingresos es en Licenciatura, seguido por la Medio superior, que representan más del 50% de nuestros ingresos.

Gráfico, Gráfico circular

Descripción generada automáticamente

La maestría y el doctorado son los niveles que tienen menos alumnos matriculados.

### Mapita:Sede

### Gráfico de barras: Programa educativo. Menor

En este grafico se puede realizar una selección por nivel y programa educativo.

Si analizamos los que tienen mayores ingresos (niveles Media Superior y Licenciatura) La Preparatoria Agrícola es la de mayor numero de alumnos, y respecto a las Licenciaturas, las liderean los de ingeniería agrónoma con las especialidades en parasitologia, agroindustrial y zootecnia.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Si realizamos un análisis de cada nivel, podemos visualizar que en el caso de doctorado es el que más demanda tiene es en ciencias en economía agrícola.

### Resumen de becados

En esta tabla se puede observar que el numero de becados externos es el mayor en comparación con otras becas que se manejan en la institución, seguido de las becas internas. Lo que se puede resaltar es que tenemos un buen programa de becas de apoyo para los alumnos.

Texto

Descripción generada automáticamente con confianza media

### Nube de palabras

En este último grafico podemos observar cuales son las nacionalidades de nuestros alumnos, los cuales la mayoría son de origen hispano siendo la mayoría mexicanos, seguido por los Colombianos, Ecuatorianos y Cubanos, por lo que la idea de realizar cursos en otros idiomas podría ser viable en el caso de que queramos tener más alumnos de otros países no hispanos, sin embargo, podemos notar que la minoría son de países en los que el español no es la lengua principal.

Texto

Descripción generada automáticamente

## 5.- Interpretación y Conclusiones.

Con respecto a nuestra primera hipótesis se descarta la idea de que el fenómeno covid fue significativo para los ingresos, por el contrario, fueron los años en los que más matriculados hubo.

Se concluye que los niveles con mayor interés son los dirigidos a la educación medio superior y licenciatura reportando 42.5% y 51.8% de los ingresos, de los cuales la mayoría de los ingresados van a las ingenierías Agrónomo especialista y Ingeniero agrónomo industrial.

El número de alumnos becados en estos dos niveles son los que muestran mayores registros.

Por lo que se debe pensar en algún programa de difusión para los niveles de maestría y doctorado, una opción puedes ser el incremento en el número de becas ya que son los niveles con menos becados y así el alumnado puede continuar con su preparación dentro de la misma institución, manteniendo la beca o incrementándola un poco para los niveles posteriores.

Por último, la mayoría del alumnado corresponde al área de Latinoamérica en donde la mayoría son Hispanohablantes, a menos que se quiera incrementar el número de alumnos de otros países como estados unidos o de algún país asiático, se puede considerar un programa en el cual la enseñanza es en inglés. Sin embargo, hasta que no se tenga difusión de ello y ver cuál sería el impacto de esta, podemos considerar no realizar ningún cambio en programa de enseñanza, ya que la mayoría pertenecen a México y países de habla hispana.