



FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
INTRODUCCIÓN A LA ANALÍTICA DE NEGOCIOS
TRABAJO DEL CURSO – PRIMERA ENTREGA: 10%
Semestre 2022-01

Equipo de trabajo: Sara Delgado, Luis Mejía, Juan Sebastián Jaramillo, Sebastián Alzate

Motivación:

Dada la gran cantidad de información que se tiene es importante encontrar escenarios de confianza que validen y analicen los datos para evidenciar tendencias, errores o simplemente justificar decisiones futuras en el marco de una organización, una ciudad o un país.

Objetivo:

Del tema que le fue asignado definir qué le interesa investigar a través de analítica descriptiva, partiendo de los conocimientos y las herramientas que aprenderá durante el curso.

Primera entrega: Para la primera entrega se deben realizar las siguientes actividades:

1. Dar un contexto de la temática:

a. ¿Qué situación se va analizar?

Se analizará la cobertura del deporte a nivel nacional, teniendo en cuenta múltiples variables que permitan hacer un sondeo tanto de la presencia a nivel nacional como el tipo de deporte que se practica en las regiones, igualmente se espera estudiar la participación a nivel internacional que se tiene en los diferentes deportes.

b. ¿Por qué le llamó la atención?

Además de aportar múltiples beneficios físicos y psicológicos, el deporte es una potente herramienta de transformación social y una actividad formadora. Es por esto que estudiar la presencia y participación en los departamentos resulta de suma importancia, pues así se puede obtener información sobre los departamentos donde hace falta una mayor participación e inversión. El deporte juega un papel importante como promotor de la integración social y el desarrollo económico en diferentes contextos geográficos, culturales y políticos; así si se estudia la representación a nivel internacional, esto puede ayudar al desarrollo del país.

c. ¿Para qué puede servir el análisis de esta situación?

Por medio de este estudio se puede explorar y conocer la presencia deportiva en el país, en los diferentes departamentos y en la variedad de deportes que existen. También se puede conseguir un contraste en el cual se observe la relación que hay de la práctica de deportes con los escenarios deportivos que hay dentro del país y si estos se encuentran en óptimas condiciones. Por último se puede plantear un análisis de la presencia internacional que tiene Colombia y revisar las posibles falencias, para tomar medidas correctivas dentro de las políticas que apoyan a este sector.

- d. Referencie mínimo 2 proyectos (pueden ser artículos o dashboards) donde se haya realizado un análisis descriptivo de la temática escogida (adjuntar referencias bibliográficas o link del dashboard). De estos explique:
- i. El objetivo: qué busca describir.
 - ii. Metodología: cómo lo hace.
 - iii. Resultados: cuáles son los resultados principales.
- **Influencia del deporte y la actividad física en el estado de salud física y mental:**
 - I. **Objetivo:** descubrir los beneficios que caracterizan la influencia del deporte y la actividad física con respecto al estado de salud tanto a nivel físico como mental.
 - II. **Método:** Mediante un proceso de estudio documental con un enfoque cualitativo y método bibliográfico. Se examinó la literatura científica incluyendo artículos de bases de datos especializadas, con participación de referencias nacionales e internacionales establecidas en campos de las ciencias del deporte, psicología, educación física y salud.
 - III. **Resultado:** El deporte y la actividad física son factores que influyen positivamente en la salud física, prevención de enfermedades crónicas o degenerativas, además de prevenir problemas de salud mental, disminuyendo tasas de nivel de estrés, depresión y ansiedad. Se encontró que el deporte mejora capacidades cognitivas, habilidades sociales, autoconcepto y resiliencia. Por estas razones el deporte genera bienestar en múltiples contextos en donde se desarrollan las personas.
 - **La participación en actividades deportivas en Colombia: enfoque empírico desde la economía, 2008-2015**
 - I. **Objetivo:** Identificar y analizar los determinantes de la decisión de participar en actividades deportivas en las cinco ciudades principales de Colombia durante el periodo 2008-2015.
 - II. **Método:** El estudio se realiza en las principales ciudades de Colombia, donde habitan casi el 50% de ciudadanos. Siguen el modelo planteado por Moffitt, donde se tienen en cuenta variables instrumentales a partir de funciones de cohorte y tiempo, se utilizan 8 secciones cruzadas y el estimador se obtiene por medio de mínimos cuadrados. Y esta función se resuelve por medio de la instrumentación en dos etapas, en la primera etapa las variables explicativas se proyectan sobre los instrumentos, en la segunda etapa se reemplazan las variables instrumentadas en el modelo inicial.
Se tiene una muestra de 48.506 hogares que contienen información de 25 cohortes para las cinco principales ciudades colombianas. En total, la estimación involucra 50 cohortes para dos grupos de participantes: asistencia a eventos deportivos (espectadores) y practicantes de actividad física y deportes.
 - III. **Resultados:** La participación de personas en actividades deportivas mantiene un comportamiento creciente hasta 2011 y a partir de allí se mantiene constante o crece poco, a excepción de Bogotá donde se observó un decrecimiento.
La ciudad con más alta participación en los ocho años es Medellín, con un poco más de 50% de participación. Cali y Cartagena presentan la menor tasa de participación. Los deportistas representativos tienen entre 18 y 25 años y cursan estudios de pregrado o posgrado y además trabajan al mismo tiempo, en su mayoría son hombres de estatus medio y alto.
La práctica y asistencia deportiva de esta población se asocia más a la satisfacción con los sitios donde se realiza. Las diferencias en la participación deportiva no solo se

derivan del lugar de residencia, el sexo y la edad, sino también del nivel educativo y de la estratificación socioeconómica, de hecho, en las cinco ciudades la participación en los estratos altos supera en 13 puntos porcentuales al de los estratos bajos. Quienes estudian, y estudian y trabajan, tienen una mayor probabilidad de realizar actividad física y de asistir a eventos deportivos. La disponibilidad de tiempo de ocio explica de mejor forma las decisiones de asistencia a eventos deportivos como espectador, que las de práctica deportiva y actividad física. El ingreso per cápita solo presenta un efecto positivo sobre la probabilidad de participar en actividades físicas y deportivas. El tipo de educación recibida por niños y jóvenes genera brechas de participación a favor de quienes estudian en la escuela privada. La pobreza es un factor determinante de la participación en estas actividades, pues el 13% de habitantes padece hambre y los objetivos son determinados por la ingesta y gasto de calorías en estas prácticas. Los lugares con mayor participación pertenecen a estratos medios y altos, donde se cuenta con infraestructura y la tasa de criminalidad y homicidios es baja.

2. Recolección y extracción: Fueron consultadas y analizadas tres bases datos relacionadas con el deporte a nivel nacional de las cuales se identificaron sus variables para posteriormente realizar consultas sobre ellas.

- **Consolidado Escenarios**

Variables:

Código: variable cualitativa nominal, la cual etiqueta cada deporte por su nombre.

Tipo: variable cualitativa nominal, la cual etiqueta el deporte por su forma de práctica.

Nombre: variable cualitativa nominal, la cual contiene el nombre de cada grupo.

Deporte: variable cualitativa nominal, la cual clasifica el tipo de deporte.

Dirección: Variable cualitativa nominal, contiene la dirección de cada escenario deportivo.

Teléfono: Variable cuantitativa discreta, contiene el número de teléfono del escenario deportivo.

Municipio: Variable cualitativa nominal, contiene el municipio de ubicación del escenario.

Departamento: Variable cualitativa nominal, contiene el nombre del departamento de ubicación del escenario.

Página web: Variable cualitativa nominal, la cual contiene la página web del ente deportivo.

Email: Variable cualitativa nominal, contiene los correos de cada ente.

Ubicación: Variable cualitativa nominal, describe el tipo de ubicación en el que se encuentra el escenario.

Razón social: Variable cualitativa nominal, contiene el nombre del ente deportivo.

Clase ubicación: Variable cualitativa nominal, describe en qué tipo de ubicación o zona se encuentra el escenario

Localidad comuna: Variable cualitativa nominal, describe el nombre de la comuna en que se encuentra ubicado el escenario deportivo.

Descripción: Variable cualitativa nominal, describe el tipo de escenario deportivo.

Latitud: Variable cuantitativa, indica la latitud de una ubicación.

Longitud: Variable cuantitativa, indica la longitud de una ubicación.

Url imagen: Variable cualitativa nominal, contiene la dirección web de una imagen del escenario.

- **Calendario Deportivo**

Variables:

Nombre: Variable cualitativa nominal, contiene el nombre del deporte.

Deporte: Variable cualitativa nominal, contiene el calendario en que se jugará

Tipo: Variable cualitativa nominal, contiene el tipo de calendario en que está programado

Fecha: Variable cuantitativa continua, contiene la fecha y hora en que se encuentra programado el encuentro.

País: Variable cualitativa nominal, contiene el país donde se realizará el evento.

Lugar: Variable cualitativa nominal, describe en que parte del país se realizará el evento.

- **Programa Escuelas Deportivas Para Todos**

Variables:

Nº: Variable cuantitativa ordinal, contiene una lista ordenada con el número de cada registro.

Departamento: Variable cualitativa nominal, contiene el nombre del departamento de ubicación de la escuela.

Municipio: Variable cualitativa nominal, contiene el nombre del municipio de ubicación de la escuela.

Deporte 1, Deporte 2, Deporte 3, Deporte 4: Son variables cualitativas nominales, describen los deportes enseñados en las escuelas.

Edad: Variable cuantitativa discreta, describe las edades de los estudiantes de la escuela.

Género: Variable cualitativa nominal, contiene el género de quienes practican deporte.

Grupo étnico: Variable cualitativa nominal, contiene el grupo étnico de quienes practican deporte.

Condición de discapacidad: Variable cualitativa nominal, esta variable hace referencia a la discapacidad que presentan las personas de la escuela.

Grupo vulnerable: Variable cualitativa nominal, describe el grupo de vulnerabilidad al que pertenecen quienes están en la escuela.

Día inscripción: Variable cuantitativa discreta, describe cuántos días llevan inscritos los estudiantes.

Mes inscripción: Variable cuantitativa discreta, describe el mes en el que se realizó la inscripción.

Año inscripción: Variable cuantitativa discreta, describe el año en que se inscribió el estudiante.

3. Carga y almacenamiento: Para esta etapa se escribe el código de desarrollo de las bases de datos seleccionadas.

- **Para Consolidado escenario:**

```
create table Consolidado_Escenarios (  
  codigo int not null,  
  tipo varchar (255) not null,  
  nombre varchar (255) not null,  
  deporte varchar (100) not null,  
  direccion varchar (255) not null,  
  telefono int null,  
  municipio varchar (255) not null,
```

departamento varchar (255) not null,
 paginaweb varchar (255) null,
 email varchar (255) null,
 ubicación varchar (255) null,
 razonsocial varchar (150) null,
 claseubicacion varchar (155) null,
 localidadcomuna varchar (155) null,
 descripcion varchar (155) null,
 latitud decimal not null,
 longitud decimal not null (decimal < 0),
 urlimagen varchar (255) not null);

- **Para Programa Escuelas Deportivas Para Todos**
 create table Programa_Escuelas_Deportivas_Para_Todos (
 N° int not null,
 Departamento varchar (150) not null,
 Municipio varchar (150) not null,
 Deporte1 varchar (150) not null,
 Deporte2 varchar (150) not null,
 Deporte3 varchar (150) not null,
 Deporte4 varchar (150) not null,
 Edad int not null,
 Género varchar (100) not null,
 Grupo_étnico varchar (100) not null,
 Condición_de_discapacidad varchar (155) not null,
 Grupo_Vulnerable varchar (155) not null,
 Día_inscripción int not null,
 Mes_inscripción int not null,
 Año_inscripción decimal not null,
 primary key (N°))

- b. Cargar las bases de datos a Bigquery. Generar 15 preguntas que le ayuden a tener un entendimiento más profundo de la información que contiene y que sean un insumo para describir el contexto de su temática. **Escribir la consulta de SQL y la salida.**

Pregunta 1: ¿Cuál es el deporte más practicado por departamento?

```
SELECT Departamento, Deporte_1, count(*) as Cuenta FROM
`entrega-analitica-n1.Datos_Deportes_Colombia.Escuelas_Deportivas`
where Departamento != 'null'
group by Departamento, Deporte_1
order by Cuenta DESC
```

Fila	Departamento	Deporte_1	Cuenta
1	CÓRDOBA	FÚTBOL	130
2	PUTUMAYO	BALONCESTO	120
3	PUTUMAYO	FÚTBOL	120
4	PUTUMAYO	FÚTBOL SALA	109
5	CÓRDOBA	FÚTBOL SALA	81
6	CÓRDOBA	BÉISBOL	80
7	PUTUMAYO	FÚTBOL DE SALÓN	80

Los deportes que más se practican en los departamentos son el fútbol y el baloncesto seguidos del fútbol sala, el béisbol y fútbol de salón.

Pregunta 2: ¿Cuántas ligas hay en el país?

```
SELECT tipo, count(*) as cuenta
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios_Deportivos`
group by tipo
```

Fila	tipo	cuenta
1	INDER	32
2	FEDERACIONES	56
3	LIGAS	599
4	CLUBES	83
5	ESCENARIO	1722

En el país actualmente se pueden encontrar 599 ligas.

Pregunta 3: ¿Qué tipo de escenario es el que más hay en el país?

```
SELECT tipo, descripcion, count(*) as Num_Esc
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios_Deportivos`
where tipo='ESCENARIO'
group by tipo, descripcion
order by Num_Esc DESC;
```

Fila	tipo	descripcion	Num_Esc
1	ESCENARIO	CANCHA MULTIPLE	526
2	ESCENARIO	CANCHA DE FUTBOL	177
3	ESCENARIO	CANCHA DE BALONCESTO	171
4	ESCENARIO	CANCHA DE MICROFUTBOL	149
5	ESCENARIO	CANCHA DE TENIS	89
6	ESCENARIO	CANCHA DE MINIFUTBOL	87

Los escenarios que más se pueden encontrar en el país son las canchas de diferentes tipos, ya sean múltiples o de un deporte en específico.

Pregunta 4: ¿Cuál es el departamento que más escenarios deportivos tiene?

```
SELECT tipo, departamento, count(*) as Numero_Esc  
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios`  
where tipo='ESCENARIO'  
group by tipo, departamento  
order by Numero_Esc desc;
```

Fila	tipo	departamento	Numero_Esc
1	ESCENARIO	VALLE DEL CAUCA	1012
2	ESCENARIO	BOLIVAR	183
3	ESCENARIO	BOYACA	100
4	ESCENARIO	ATLANTICO	100
5	ESCENARIO	RISARALDA	95
6	ESCENARIO	CUNDINAMARCA	86
7	ESCENARIO	CORDOBA	70
8	ESCENARIO	CALDAS	37

El departamento que cuenta con mayor número de escenarios deportivos en el Valle del Cauca, el cual sobrepasa por mucho los demás departamentos.

Pregunta 5: ¿Cuál es el escenario deportivo que más hay en el Valle del Cauca?

```
SELECT tipo, departamento, descripcion, count(*) as Numero_Esc  
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios`  
where tipo='ESCENARIO'  
group by tipo, departamento, descripcion  
order by Numero_Esc desc;
```

Fila	tipo	departamento	descripcion	Numero_Esc
1	ESCENARIO	VALLE DEL CAUCA	CANCHA MULTIPLE	227
2	ESCENARIO	VALLE DEL CAUCA	CANCHA DE BALONCESTO	130
3	ESCENARIO	VALLE DEL CAUCA	CANCHA DE FUTBOL	115
4	ESCENARIO	VALLE DEL CAUCA	CANCHA DE MICROFUTBOL	104

En el Valle del Cauca, se encuentran muchos escenarios deportivos comprendidos principalmente por canchas múltiples y de baloncesto.

Pregunta 6: ¿Qué deporte es el que más escenarios tiene?

```
SELECT deporte, count(*) as Numero_Esc
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios`
where deporte!= 'null'
group by deporte
order by Numero_Esc desc;
```

5	FUTBOL	68
6	CICLISMO	29
7	ATLETISMO	28

El deporte que cuenta con mayor número de escenarios para practicarse es el Fútbol, seguido por el ciclismo y el atletismo, lo cual se puede deber principalmente al hecho de que son deportes en los cuales el país ha recibido reconocimientos.

Pregunta 7: ¿Qué departamento tiene el mayor número de niños en algún deporte?

```
SELECT replace(replace(replace(lower(Deporte_1),'ú','u'),'ó','o'),' ','') as deporte,
       replace(replace(replace(replace(lower(Departamento),'í','i'),'ó','o'),'á','a'),' ','') as departamento,
       count(*) as numero
FROM `primera-entrega-355623.Deporte.Escuelas`
where Deporte_1 != 'null'
group by deporte, departamento
order by numero desc;
```

Fila	deporte	departamento	numero
1	futbol	bolivar	1288
2	futbol	choco	1184
3	futbol	santander	727
4	futbol	cauca	684
5	futbol desalon	tolima	636
6	atletismo	antioquia	605
7	futbol	nariño	530

Los niños tienen una mayor participación en deportes en los departamentos como Bolívar, Chocó y Santander.

Pregunta 8: Edad promedio de los niños que practican determinado deporte

```
select replace(replace(replace(replace(lower(Deporte_1),'ú','u'),'ó','o'),'é','e'),' ','') as deporte,
       round(avg(Edad),2) as edad_prom
FROM `primera-entrega-355623.Deporte.Escuelas`
where Deporte_1 != 'null'
group by deporte
```


order by deporte;

Fila	deporte	edad_prom
1	ajedrez	9.46
2	atletismo	9.89
3	baloncesto	9.99
4	balonmano	10.35
5	balonpesado	9.29
6	beisbol	9.7
7	boxeo	9.68
8	canotaje	9.28
9	ciclismo	10.16
10	ciclomontañismo	9.76

La edad promedio de los niños en cada deporte es cerca de los 10 años.

Pregunta 9: ¿Cuántos estudiantes hay por deporte y por sexo?

```
SELECT replace(replace(replace(replace(lower(Deporte_1),'ú','u'),'ó','o'),'é','e'),' ','') as deporte,  
        lower(G__nero) as sexo, count(*) as numero_Estudiantes  
FROM `primera-entrega-355623.Deporte.Escuelas`  
where G__nero != 'null'  
group by deporte,sexo  
order by numero_Estudiantes desc;
```

Fila	deporte	sexo	numero_Estudiantes
1	futbol	masculino	8034
2	futboldesalon	masculino	3866
3	baloncesto	femenino	1944
4	baloncesto	masculino	1806
5	futbol	femenino	1710
6	futboldesalon	femenino	1498
7	voleibol	femenino	1345
8	voleibol	masculino	1104

Pregunta 10: ¿Cuál es la edad promedio de los estudiantes que practican deporte en los departamentos?

```
SELECT replace(replace(replace(replace(lower(Departamento),'í','i'),'ó','o'),'á','a'),' ','') as  
        departamento, round(avg(Edad),2) as edad_prom  
FROM `primera-entrega-355623.Deporte.Escuelas`  
where Departamento != 'Año inscripción'  
group by departamento  
order by edad_prom desc;
```


Fila	etnia	cuenta
1	afro	6181
2	ninguna	20705
3	indigena	2900
4	campesino	1165
5	raizal	888
6	rom	160
7	palenquero	74
8	mestizo	187

Los estudiantes que asisten mayormente a este tipo de escenarios deportivos no identifican su etnia, sin embargo la etnia que presenta mayores integrantes son la afro e indigena.

Pregunta 14: ¿Cuál es el deporte con más clubes en el país?

```
SELECT tipo, deporte, count(*) as Numero_Clubes
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios`
where tipo= 'CLUBES'
group by tipo, deporte
order by Numero_Clubes desc;
```

Fila	tipo	deporte	Numero_Clubes
1	CLUBES	FUTBOL	36
2	CLUBES	PARAPENTE	20
3	CLUBES	AUTOMOVILISMO	13
4	CLUBES	PARACAIDISMO	7
5	CLUBES	ALA DELTA	3
6	CLUBES	AEROMODELISMO	1

Pregunta 15: ¿Cuál es el departamento con más federaciones?

```
SELECT tipo, departamento, count(*) as Numero_Federaciones
FROM `entrega-analitica-n1.Datos_Deportes_Colombia.Escenarios`
where tipo= 'FEDERACIONES'
group by tipo, departamento
order by Numero_Federaciones desc;
```

Fila	tipo	departamento	Numero_Federaciones
1	FEDERACIONES	BOGOTÁ D.C.	42
2	FEDERACIONES	VALLE DEL CAUCA	5
3	FEDERACIONES	ANTIOQUIA	3
4	FEDERACIONES	BOLIVAR	2
5	FEDERACIONES	NARIÑO	1
6	FEDERACIONES	TOLIMA	1
7	FEDERACIONES	ATLANTICO	1

Referencias:

<https://dialnet.unirioja.es/servlet/articulo?codigo=6369972>
<https://revistas.urosario.edu.co/index.php/economia/article/view/10472>
<https://revistas.um.es/sportk/article/view/461671>
<https://www.datos.gov.co/Mapas-Nacionales/Consolidado-Escenarios/886w-94ug>
<https://www.datos.gov.co/Deporte-y-Recreaci-n/Programa-Escuelas-Deportivas-Para-Todos/4e7j-65ci>
<https://www.datos.gov.co/Deporte-y-Recreaci-n/Calendario-Deportivo/m5xp-39nw>
https://www.scielo.cl/scielo.php?pid=S0718-07642021000300149&script=sci_arttext&tlng=pt