

DUOC UC
2024

PROYECTO DATALAB DUOC

Integrantes: Patricia Aldunce
Cindy Cañete
Francisco González



CONTENIDO

- Resumen de avance & Objetivo
- Metodología
- Evidencia de desarrollo
- Ajustes
- Planificación

RESUMEN DE AVANCE

OBJETIVO

- Carga y Verificación de Datos
- Identificación de Variables Clave
- Revisión de la Calidad de los Datasets
- Enfoque en Variables con Mejores Insights
- Preparación para Análisis para Fase 3

Se mantienen los objetivos establecidos en el documento anterior titulado '1.5_GuiaEstudiante_Fase 1_Definicion Proyecto APT'



METODOLOGÍA

- **Metodología ágil**

El proyecto cuenta con un backlog detallado de las próximas historias a desarrollar, considerando el proyecto finalizado

- **Seguimiento semanal**

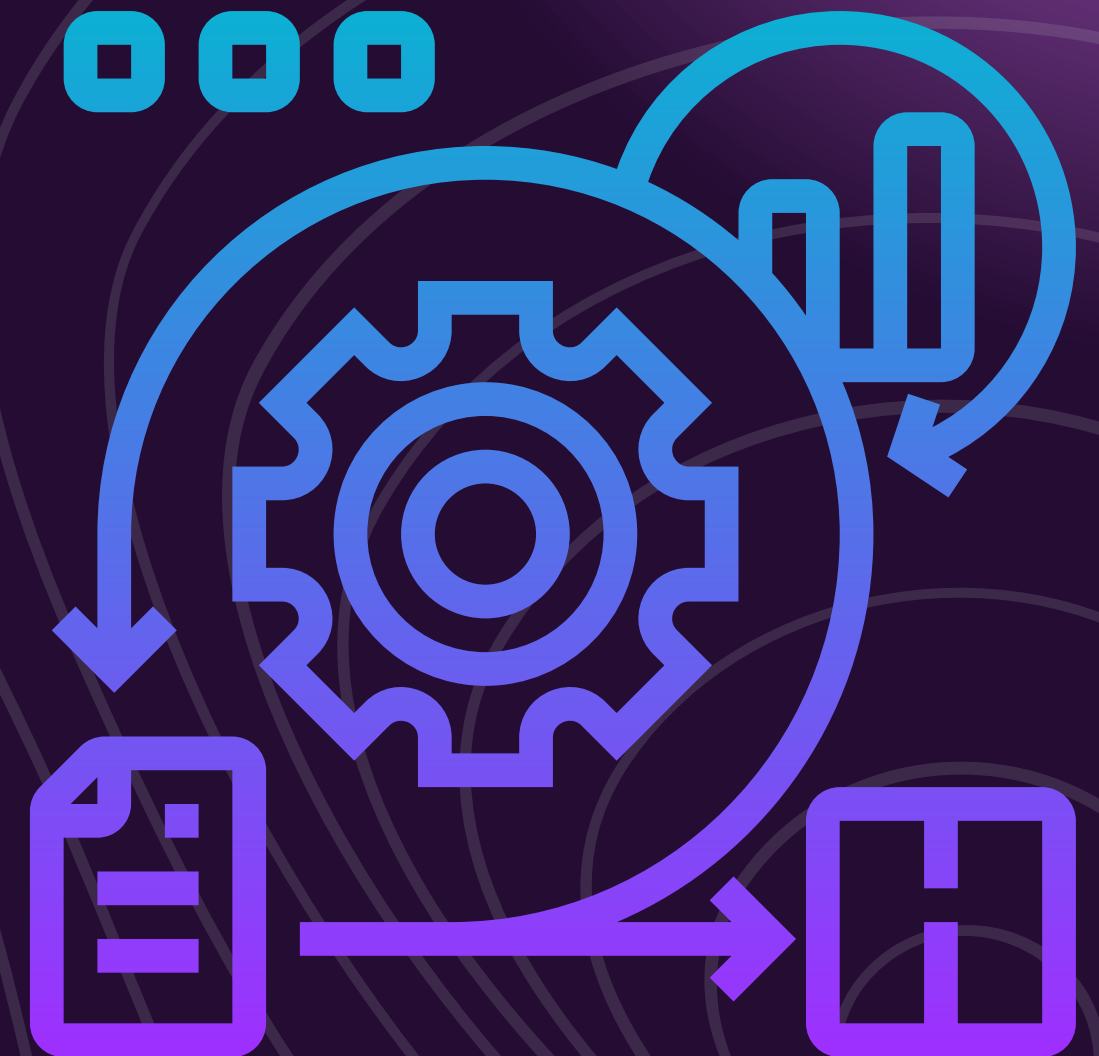
Reuniones todos los domingos de 11:15 a 12:15 se refina las HDU tanto técnicas como de negocio. Además, los lunes de 20:00 a 21:00 nos reunimos para discutir el estado de la semana, abordando aspectos como enfoques, bloqueos o cualquier impedimento en nuestras funciones

- **Canales oficiales de comunicación**

1. Whatsapp: Para bloqueos de desarrollo avisar de inmediato al grupo
2. Discord: Para salas de reuniones online, ya sea registrada en el calendario o esporádico

- **Registro de avance de proyecto**

Jira: Para las historias de usuarios y/o incidentes





EVIDENCIA DE AVANCE DE FASE 1 Y 2 DE LA METODOLOGÍA CRISP-DM



PRÁCTICAS APLICADAS

```
data = pd.read_csv("ofertas-5años-duoc.csv")
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366306 entries, 0 to 366305
Data columns (total 40 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Id oferta              365806 non-null float64
1   Oferta                 365789 non-null object
2   Creación               365806 non-null object
3   Publicación            365806 non-null object
4   Estado                 365806 non-null object
5   Término esperado       365758 non-null object
6   Término                 360333 non-null object
7   Id Empresa              365806 non-null float64
8   RUT                    355722 non-null object
9   Empresa                365800 non-null object
10  Tipo Empresa            365806 non-null object
11  Categorías empresa      41253 non-null object
12  Rubro                   361969 non-null object
13  Contacto                365657 non-null object
14  Contacto fono           347253 non-null object
15  Contacto email          365806 non-null object
16  Contacto cargo          237669 non-null object
17  Área de trabajo        363637 non-null object
18  Tipo de cargo           363483 non-null object
19  Carrera(s)              348175 non-null object
20  Idioma(s)               26360 non-null object
21  Modalidad               365806 non-null object
22  País                    332564 non-null object
23  Región                  332564 non-null object
24  Comuna                  332563 non-null object
25  Inclusiva                365806 non-null object
26  Facultad                347793 non-null object
27  Sueldo                  363552 non-null float64
28  Postulantes             365806 non-null float64
29  Tipo de oferta          365806 non-null object
30  Jornada                 365806 non-null object
31  Vacantes                365806 non-null float64
32  Visitas                 365806 non-null float64
33  Nº Invitaciones         365806 non-null float64
34  Seleccionó               365806 non-null object
35  N Colocados             365806 non-null float64
36  Exp min                 365806 non-null float64
37  Exp máx                 365806 non-null float64
38  # repubicaciones        365806 non-null float64
39  Link                    365806 non-null object
dtypes: float64(11), object(29)
memory usage: 111.8+ MB
```

```
data.isnull().sum()

Id oferta              500
Oferta                 517
Creación               500
Publicación            500
Estado                 500
Término esperado       548
Término                 5973
Id Empresa              500
RUT                    10584
Empresa                506
Tipo Empresa            500
Categorías empresa     325053
Rubro                   4337
Contacto                649
Contacto fono          19053
Contacto email          500
Contacto cargo         128637
Área de trabajo        2669
Tipo de cargo           2823
Carrera(s)             18131
Idioma(s)               339946
Modalidad               500
País                    33742
Región                  33742
Comuna                  33743
Inclusiva                500
Facultad                18513
Sueldo                  2754
Postulantes             500
Tipo de oferta          500
Jornada                 500
Vacantes                500
Visitas                 500
Nº Invitaciones         500
Seleccionó              500
N Colocados             500
Exp min                 500
Exp máx                 500
# repubicaciones        500
Link                    500
dtype: int64
```

data.describe()											
	Id oferta	Id Empresa	Sueldo	Postulantes	Vacantes	Visitas	Nº Invitaciones	N Colocados	Exp min	Exp máx	republica
count	365806.000000	365806.000000	3.635520e+05	365806.000000	365806.000000	365806.000000	365806.000000	365806.000000	365806.000000	365806.000000	365806.000000
mean	291163.609615	180655.586505	1.065221e+13	16.528439	3.720748	280.655746	0.004265	0.063381	0.672482	10.615326	2.000000
std	158326.139449	123384.248705	3.216265e+15	25.188066	8.062890	307.856404	0.602749	0.673751	1.360114	8.696856	7.000000
min	9416.000000	1.000000	0.000000e+00	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	2.000000	0.000000
25%	166454.000000	80692.000000	1.500000e+05	3.000000	1.000000	98.000000	0.000000	0.000000	0.000000	2.000000	0.000000
50%	255033.000000	178745.000000	4.000000e+05	8.000000	1.000000	202.000000	0.000000	0.000000	0.000000	5.000000	0.000000
75%	393479.000000	275489.000000	6.900000e+05	20.000000	3.000000	367.000000	0.000000	0.000000	2.000000	21.000000	1.000000
max	628397.000000	527084.000000	1.111111e+18	395.000000	600.000000	7152.000000	150.000000	66.000000	10.000000	21.000000	135.000000

COLUMNAS CLAVES REPORTE OFERTAS

Posibles columnas clave

- 1. Oferta: Descripción de la práctica o puesto.
- 2. Categorías empresa: Podría reflejar el tipo de empresa, útil para identificar áreas de especialización.
- 3. Área de trabajo: Especifica el área de especialización de la práctica.
- 4. Tipo de cargo: Detalla el tipo de rol ofrecido, lo que puede relacionarse con el tipo de práctica.
- 5. Modalidad: Indica si la práctica es presencial o remota, un factor clave de segmentación.
- 6. Región y Comuna: Representan la ubicación geográfica, importante para segmentar las ofertas por región.
- 7. Inclusiva: Indica si la oferta es inclusiva para personas con discapacidad, lo cual es clave para fomentar equidad.
- 8. Sueldo: El salario ofrecido para la práctica, importante para segmentar por nivel económico.
- 9. Exp min y Exp máx: Los años de experiencia requeridos, para clasificar según la preparación de los postulantes.
- 10. Vacantes: Número de puestos disponibles, útil para entender la demanda.
- 11. Postulantes: Número de postulantes, para evaluar la competitividad de la oferta.

COLUMNAS CLAVES REPORTE POSTULACIONES

- 1. Segmentación y análisis de patrones (clustering)

Para segmentar y encontrar patrones en las postulaciones y ofertas laborales, serán útiles las variables relacionadas con:

Id Empresa / Tipo empresa / Rubro: Para identificar el tipo de empresas que publican las ofertas. **Tipo oferta / Fecha publicación / Sueldo ofrecido:** Ayudarán a segmentar las ofertas según características comunes. **Experiencia Mínima y Máxima Requerida:** Agruparán ofertas por el nivel de experiencia solicitado. **Estado postulación:** Para ver qué ofertas tienen más éxito y cuáles no logran atraer postulantes.

- 2. Creación de modelos predictivos

Para predecir el comportamiento de las vacantes y postulaciones:

Fecha publicación / Fecha postulación: Para analizar el tiempo que toma una oferta en recibir postulaciones. **Sueldo ofrecido / Expectativa de sueldo:** Estas variables serán esenciales para predecir qué tipo de oferta genera más interés en los postulantes. **Estado postulación / Entrevista / Estado entrevista:** Puede predecir la probabilidad de que una postulación pase a entrevista o sea exitosa.

- 3. Clasificación de oferta y perfiles de usuarios

Para clasificar y construir perfiles de los postulantes:

Id usuario: Variable identificativa, aunque no es necesaria para el análisis, es importante para el manejo de datos. **Experiencia / Fecha de nacimiento:** Pueden ser útiles para clasificar a los postulantes según su nivel de experiencia y edad. **Programa 1 candidato / Facultad 1 / Sede 1:** Permitirán clasificar postulantes según la formación académica.

- 4. Visualización y reportes dinámicos

Todas las variables numéricas (**Sueldo ofrecido, Expectativa de sueldo, Duración entrevista, Experiencia**) son ideales para gráficos y análisis visual. **Categorías / Rubro / Estado postulación:** Variables categóricas que serán útiles para crear filtros interactivos en Power BI.

COLUMNAS CLAVES REPORTE USUARIOS

VARIABLES RELEVANTES

1. Sexo: Indica categorías femenino o masculino. Ver como alternativa para hacer análisis de género.
2. Fecha de nacimiento: Para calcular la edad de los usuarios que actualizan sus registros.
3. Estado civil: Indica estado civil de usuario, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
4. País: Variable geográfica, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
5. Región: Variable geográfica, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
6. Comuna: Variable geográfica, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
7. Programa: Variable que indica carrera del usuario (estudiante o titulado).
8. Estado: Variable que indica si el usuario es estudiante o titulado, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
9. Facultad: Variable que indica la Escuela a la que pertenece el estudiante o titulado, relevante para hacer segmentación o perfilamiento de los usuarios que utilizan la plataforma.
10. Años de experiencia: Variable que indica años de experiencia.
11. Estado laboral: Variable binaria que indica si el usuario busca o no.
12. Actualización: Variable que indica la fecha en que el usuario realizó una actualización de su perfil. Esta variable puede ayudar a determinar que usuarios se encuentran más activos o la frecuencia con la que utilizan el portal.
13. Completitud: Indica el porcentaje de completitud que tiene el usuario respecto de su perfil en la plataforma. Puede ser relevante para hacer algún tipo de clusterización o perfilamiento.
14. Buscando trabajo: Variable binaria, que indica si el usuario se encuentra buscando trabajo o no.
15. Fecha actualización de preferencias: Variable que indica la fecha en que el usuario realizó una actualización las preferencias de búsqueda tipo de empleo. Esta variable puede ayudar a determinar que usuarios se encuentran más activos o la frecuencia con la que utilizan el portal.
16. Tiene alguna discapacidad: Indica si el usuario tiene alguna discapacidad o no, este campo es declarativo. Relevante para realizar análisis respecto a temáticas de inclusión.

AJUSTES AL MODELO PREDICTIVO

Se realizó un leve ajuste en el segundo punto del alcance, añadiendo una especificación: el modelo se centrará en prever cuántas empresas estarán publicando ofertas de prácticas en 2025 y cómo esta proyección puede impactar las oportunidades disponibles para los futuros practicantes.

AJUSTES AL MODELO PREDICTIVO		
Aspecto	Modelo Anterior	Modelo Actual
Foco del análisis	Predicción del comportamiento del mercado laboral en relación con vacantes y prácticas profesionales.	Predicción de la oferta laboral de las empresas para el año 2025
Objetivo predictivo	Anticipar la demanda de prácticas profesionales para ajustar la oferta educativa.	Predecir cómo evolucionará la oferta laboral de las empresas para el 2025 .
Usuarios principales	Escuelas de Duoc UC y otros actores institucionales para adaptar sus programas educativos.	
Enfoque temporal	Enfocado en analizar el comportamiento actual y predecir la demanda de prácticas en el corto plazo.	Centrado en proyecciones futuras a mediano plazo (2025) basadas en el comportamiento histórico.
Impacto esperado	Mejora de la oferta educativa alineada con las tendencias del mercado laboral.	Mejora en la planificación de recursos y estrategias de las escuelas para satisfacer la demanda laboral futura.

CUMPLIENDO HITO DE LA PLANIFICACIÓN

La planificación indicada al inicio del proyecto se mantiene sin cambios

Semana 8 finalizada con éxito

[illegible]



**EL ANÁLISIS CON MACHINE
LEARNING NO ES UN PRODUCTO
FINAL; ES UN PROCESO
CONTINUO DE MEJORA Y
REFINAMIENTO**

Muchas gracias por la atención.