Estudiante	MARIA CECILIA ZUBIN
Curso/Comisión	61615
Proyecto	Data set / Análisis de Variables / Planteo de Hipótesis – Preentrega – 20/08/2024

Temática a Resolver

En base a datos históricos de un trimestre, una empresa de servicios nos consulta con el objeto de predecir e inferir si el título universitario es un elemento fundamental al momento de acceder a puestos jerárquicos percibiendo un salario superior promedio más alto que las personas que ni tiene título universitario.

Adicionalmente, desean saber si la ubicación geográfica, en donde la empresa tiene establecimientos, está vinculada o no a la obtención de un título universitario completo.

Como tercera alternativa resulta interesante, analizar de la población que posee título universitario, que proporción pertenecen al sexo masculino y cuanta al sexo femenino.

La intención de la empresa con esta solicitud, es implementar un sistemas interno a través del departamento de desarrollo de RRHH un esquema de beneficios & compensaciones especialmente direccionado a promover e incentivar al personal a estudiar una carrera universitaria y/o finalizarla en los casos en los que se encuentre incompleta.

Resumen de Hipótesis

Hipótesis Nº1: El poseer un titulo universitario es un factor determinante para tener acceso a puestos jerárquicos, por ende, de mayor ingreso.

Hipótesis №2: La ubicación geográfica del empleado influye en la obtención de un título universitario.

Hipótesis №3: La población de sexo femenino que posee título universitario es mayor, que la población de sexo masculino,

Detalle de variables contempladas en el dataset

Legajo: numero identificador del empleado, ID univoco que no se repite dentro del dataset.

Estado: hace referencia a si el empleado se encuentra activo al momento del armado del dataset.

Estructura: detalla la estructura de liquidación bajo la cual se encuentra contenido el empleado. Los puestos asociados a staff, analistas, supervisores, responsables, etc., son los asociados al personal fuera de convenio. El personal que realiza tareas operativas detalladas explícitamente en una convención colectiva de trabajo nomenclada y registrada como tal, están asociados al personal dentro de convenio. [Fuera de convenio, Dentro de convenio]

Convenio: detalla los posibles convenios al que puede pertenecer el empleado [FCON,CC-781,ATCC]

ID_Convenio	Convenio del Empleado	
33	FCON	
22	CC-781	
11	ATCC	

Categoría: detalla la categoría a la que pertenece el empleado:

20,0 Hs. CATEGORIA 3 CCT688/14

ID_Cat	Categoría	ID_Cat	Categoría
FCC	Fuera de Convenio	OP20	20 Hs. CATEGORIA OP. A CCT 781
H01	36,0 Hs. CATEGORIA 3 CCT688/14	OP24	24 Hs. CATEGORIA OP. A CCT 781
H03	35,0 Hs. CATEGORIA 3 CCT688/14	OP25	25 Hs. CATEGORIA OP. A CCT 781
H11	31,0 Hs. CATEGORIA 3 CCT688/14	OP27	27 Hs. CATEGORIA OP. A CCT 781
H13	30,0 Hs. CATEGORIA 3 CCT688/14	OP30	30 Hs. CATEGORIA OP. A CCT 781
H21	26,0 Hs. CATEGORIA 3 CCT688/14	OP35	35 Hs. CATEGORIA OP. A CCT 781
H23	25,0 Hs. CATEGORIA 3 CCT688/14	OP36	36 Hs. CATEGORIA OP. A CCT 781
H25	24,0 Hs. CATEGORIA 3 CCT688/14		
H30	21,5 Hs. CATEGORIA 3 CCT688/14		
i			

ID_Cat: código de la categoría del empleado.

Ingreso: fecha de ingreso del empleado a la empresa.

Sexo: sexo del empleado [M: Hombre, F: Mujer]

FOR

Formador

Provincia: sucursal en la que el empleado se encuentra actualmente prestando servicios [FORM,CBA, CABA2]

Puesto: se detallan las codificaciones y puestos de todo el personal

ID_PuestoBasico	Puesto	ID_PuestoBasico	Puesto
ACC	Analista de Calidad & Capacit.	GAF	Gerente de Administración & F
ACE	Analista de Customer Experience	GCG	Gerente Ctrol. Gestion
ACH	Analista Gestion Cap. Humano	GCM	Gerente de Comunicaciones y Marketing
ACI	Analista de Comunicaciones	GCO	Gerente Comercial
ACS	Analista de Control C de Serv	GEO	Gerente Operaciones
ADP	Analista de Adminis. de Pers.	GLE	Gerente de Legales
ADS	Analista de Servidores	GMI	Gerente Mantenim. e Infraest.
AFR	Analista de Fraude	GRH	Gerente de RRHH
AGE	Analista de Gestion	GST	Gerente de Sistemas & Tecnolog
AGS	Analista de Gestión Senior	JAN	Jefe Administracion de Personal y Novedades
AMA	Analista de Mesa de Ayuda	JCR	Jefe de Comunicaciones & RSE
AMG	Analista de Modelo de Gestión	JDC	Jefe Comercial
AMT	Analista de Mantenimiento	JDS	Jefe de Desarrollo de Sistemas
ANL	Analista	JFD	Jefe Formacion & Desarrollo
ANR	Analista de Novedades y RRHH	JGP	Jefe de Contrrol de Gestión y PlanOperativo
APA	Analista de Aplicaciones	JNS	Jefe de NOC y Seguridad de la Información
APO	Analista de Planeamiento Oper.	JOP	Jefe de Operaciones
ARF	Analista de Rep. & Facturacion	JRM	Jefe RRLL & Medic. Laboral
ARH	Analista de Recursos Humanos	JSH	Jefe Seguridad e Higiene
ARL	Analista de Relaciones Labor.	LDT	Lider de Telefonia
ASE	Analista de Seleccion	REC	Recepcionista
AST	Analista de Soporte Tecnico	RSG	Responsable Servicios Grales.
ASY	Analista de Selección y Desarrollo	SCA	Supervisor de Calidad
CAL	Analista de Calidad	SCC	Supervisor Calidad & Capacit.
CAP	Coordinador Adm. de Personal	SCL	Supervisor de Operaciones
CDT	Operador	SEO	Analista Soporte Estratégico Operativo
CFA	Coordinador de Facturación	SIM	Supervisor de Infraestructura & Mantenimiento
CGC	Coordinador de Gestión Comercial	SPO	Supervisor de Plan. Operativo
CLW	Coordinador de LAN / WAN	SPR	Supervisor de Reporting
COO	Coordinador	SUN	Supervisor de Novedades
СРА	Analista de Compras	SWF	Supervisor de Workforce
СРО	Coordinador Planea. Operativo	TLC	Team Leader Operaciones
CSE	Coordinador de Selección & Emp		
CSS	Coordinador de Servicios IT		
CST	Coordinador de Soporte Tecnico		
DIE	Director Ejecutivo		
DIO	Director de Operaciones		
EJU	Ejecutivo Comercial		

Pirámide: detalla el nivel de jerarquía en el organigrama de la empresa, siendo 8 el valor que se ubica en la base de la pirámide y 0 el ultimo valor de la cima de la misma [0,1,2,3,4,5,6,7,8]

Título: hace referencia a la posesión del mismo [SI,NO]

ID_PuestoBasico: código del puesto del empleado.

ID Banco: código del banco donde el empleado percibe sus haberes.

Banco: detalle de banco donde el empleado percibe sus haberes:

ID_Banco	Banco Pago de Rem		
7	BANCO DE GALICIA Y BUENOS AIRES S.A.		
11	BANCO DE LA NACION ARGENTINA		
15	INDUSTRIAL AND COMMERCIAL BANK OF CHINA		
17	BBVA BANCO FRANCES S.A.		
20	BANCO DE LA PROVINCIA DE CORDOBA S.A.		
27	BANCO SUPERVIELLE S.A.		
44	BANCO HIPOTECARIO S.A.		
72	BANCO SANTANDER RIO S.A.		
150	HSBC BANK ARGENTINA S.A.		

Análisis Exploratorio de Datos (EDA)

El EDA se centrará en:

- Resumen estadístico de las variables clave.
- Visualización de distribuciones de ingreso, jerarquía de puesto y título universitario.

```
Legajo
                      object
Estado
                      object
Estructura
                      object
Convenio
                      object
Categoria
                      object
ID_Cat
                      object
Ingreso
                      object
Sexo
                      object
provincia
                      object
                      object
puesto
piramide
                       int64
Titulo
                      object
ID_PuestoBasico
                      object
ID_Banco
                       int64
Banco Pago de Rem
                      object
Enero
                       int64
                       int64
Febrero
Marzo
                       int64
dtype: object
```

Se detalla el tipo de cada variable.

Estructura del dataset

```
[150] df.shape

(2205, 18)

El dataset obtenido posee 2205 filas y 18 columnas
```

Análisis de existencia de valores duplicados.



Análisis de existencia de posibles valores nulos.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2205 entries, 0 to 2204
Data columns (total 18 columns):
#
    Column
                     Non-Null Count Dtype
0
   Legajo
                     2205 non-null object
1
   Estado
                     2205 non-null object
2
                    2205 non-null
                                    object
   Estructura
   Convenio
                     2205 non-null object
4
   Categoria
                    2205 non-null object
   ID_Cat
                     2205 non-null object
6
   Ingreso
                     2205 non-null object
   Sexo
                     2205 non-null object
                    2205 non-null
8
    provincia
                                    object
9
    puesto
                    2205 non-null object
10 piramide
                    2205 non-null int64
11 Titulo
                    2205 non-null
                                    object
12 ID_PuestoBasico 2205 non-null object
13 ID Banco
                     2205 non-null int64
14 Banco Pago de Rem 2205 non-null
                                    object
                      2205 non-null
                                    int64
15 Enero
16 Febrero
                     2205 non-null
                                    int64
                      2205 non-null
                                    int64
17 Marzo
dtypes: int64(5), object(13)
memory usage: 310.2+ KB
```

Propuesta de Visualizaciones para la primera etapa del trabajo

- 1. Boxplot
- 2. Histograma
- 3. Pie Chart

Primeras conclusiones

El presente trabajo fue realizado bajo el concepto de estudio solicitado inicialmente, realizando una única iteración sobre las variables elegidas, los motivos y la escueta selección de modelos.

En esta primera instancia, las hipótesis se cumplieron.