

Nombre	Código	Rol 1	Rol 2
Luisa Vélez	202022342	Análisis de datos	Despliegue
Nicolas Álvarez	202020565	Ingeniería de datos	Tablero de Datos
Luis Cortés	201924904	Análisis de negocio	Ciencia de Datos

## Reporte Proyecto 1

Link repositorio de GitHub: <https://github.com/NicoAO/RepositorioNico>

**Cliente de interés:** El área de producción de la empresa, interesada especialmente en los tiempos y tasas de producción.

### Tarea 1 (Análisis de Negocio)

Considerando el interés del área de producción de la empresa, se plantearon las siguientes preguntas:

- ¿Cómo es el comportamiento de los datos obtenidos?
- ¿Qué factores están afectando positivamente la productividad actual de la empresa? ¿en qué medida?
- ¿Qué factores están afectando negativamente la productividad actual de la empresa? ¿en qué medida?

Para poder responder estas preguntas, se pueden realizar gráficos descriptivos de los datos obtenidos (gráficos de cajas y bigotes, histogramas, gráficos de violín) con el fin de entender el comportamiento de la empresa. Por otro lado, se pueden realizar comparaciones entre las variables con gráficos de dispersión y una matriz de correlación, para poder entender la relación entre las variables, específicamente con la producción actual. De esta manera, se podrá visualizar cuales variables están afectando positivamente y cuales negativamente la productividad. Finalmente, se puede realizar un modelo predictivo a partir de una regresión lineal, capaz de estimar la productividad, teniendo en cuenta los parámetros ingresados por el usuario. Este modelo ayudaría a medir de mejor manera que tanto está afectando cada variable a la producción.

### Tarea 2 – Limpieza y alistamiento de datos (Ingeniería de datos)

La limpieza de datos es el proceso de arreglar o eliminar datos incorrectos, duplicados o incorrectos en el dataset. De acuerdo con el libro de Harvard Business Review, “Data analytics for managers”, uno de los primeros pasos es eliminar duplicados y observaciones irrelevantes, luego sigue arreglar errores estructurales como los datos faltantes (NaN) y finalmente identificar los outliers para entender el efecto que podrían tener en el modelo y que relaciones hay entre ellos.

Por lo tanto, procedemos a buscar **duplicados** y encontramos que no hay filas duplicadas. Luego, buscamos si hay **valores faltantes NaN** y posteriormente encontramos que hay 506 NaNs en la columna WIP, los datos faltantes corresponden al departamento de “Finishing”, al haber definido al cliente como el área de producción, se consideró relevante mantener la variable de wip y **separar los datos por departamentos** de forma que para el departamento de *sweing* se mantiene la columna wip y en el departamento de *Finishing* se elimina dicha columna para eliminar los datos faltantes.

Departamento Sewing

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time
0	1/1/2015	Quarter1	sweing	Thursday	8	0.80	26.16	1108.0	7080
2	1/1/2015	Quarter1	sweing	Thursday	11	0.80	11.41	968.0	3660
3	1/1/2015	Quarter1	sweing	Thursday	12	0.80	11.41	968.0	3660
4	1/1/2015	Quarter1	sweing	Thursday	6	0.80	25.90	1170.0	1920
5	1/1/2015	Quarter1	sweing	Thursday	7	0.80	25.90	984.0	6720
...	...	...	...	...	...	...	...	...	...
1187	3/11/2015	Quarter2	sweing	Wednesday	4	0.75	26.82	1054.0	7080
1188	3/11/2015	Quarter2	sweing	Wednesday	5	0.70	26.82	992.0	6960
1189	3/11/2015	Quarter2	sweing	Wednesday	8	0.70	30.48	914.0	6840
1190	3/11/2015	Quarter2	sweing	Wednesday	6	0.70	23.41	1128.0	4560
1191	3/11/2015	Quarter2	sweing	Wednesday	7	0.65	30.48	935.0	6840

Departamento Finishing

	date	quarter	department	day	team	targeted_productivity	smv	over_time	incentive
1	1/1/2015	Quarter1	finishing	Thursday	11	0.70	4.15	1440	0
2	2/4/2015	Quarter1	finishing	Wednesday	2	0.80	3.94	2160	0
2	2/4/2015	Quarter1	finishing	Wednesday	1	0.80	3.94	1200	0
2	2/4/2015	Quarter1	finishing	Wednesday	8	0.70	4.15	3000	0
2	2/4/2015	Quarter1	finishing	Wednesday	7	0.70	3.94	960	0
...	...	...	...	...	...	...	...	...	...
3	3/11/2015	Quarter2	finishing	Wednesday	10	0.75	2.90	960	0
3	3/11/2015	Quarter2	finishing	Wednesday	8	0.70	3.90	960	0
3	3/11/2015	Quarter2	finishing	Wednesday	7	0.65	3.90	960	0
3	3/11/2015	Quarter2	finishing	Wednesday	9	0.75	2.90	1800	0
3	3/11/2015	Quarter2	finishing	Wednesday	6	0.70	2.90	720	0

```

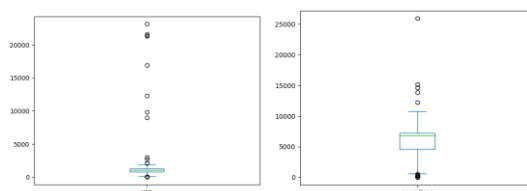
NaNs por cada fila del DataFrame:
over_time      0
incentive      0
date           0
idle_time      0
quarter        0
idle_men       0
department     0
day            0
no_of_style_change  0
team          0
no_of_workers  0
targeted_productivity  0
actual_productivity  0
smv            0
wip            0
dtype: int64

```

Posteriormente para identificar los **outliers**, seleccionamos las columnas numéricas y tomando como base el teorema del límite central, utilizamos la herramienta de los **z-scores**, el z-score es una herramienta que entrega una puntuación que indica a cuántas desviaciones estándar por encima o por debajo de la media se ubica un valor. Un z-score mayor a 3, indica que los valores están 3 desviaciones por encima de la media, lo que caracterizaría anomalías en la observación.

Por lo tanto, calculamos los *valores z* y creamos un dataframe llamado “outliers” en el que se incluyen las observaciones anormales, éste dataframe tiene 97 filas de observaciones donde alguna columna numérica tiene un z-score mayor a 3.

$$z = np.abs(stats.zscore(array\ de\ columnas\ numéricas)) \rightarrow Outliers = DataFrame[z > 3]$$



Finalmente, se escribe un algoritmo para identificar en cuáles equipos se presenta mayor WIP o mayor *overtime* que de acuerdo con los *boxplots*, son las variables que presentan mayor amplitud de caja y asimismo, mayor número de outliers por encima de los “bigotes” de la caja. El equipo que más aparece en los outliers es el equipo 11 con 20 observaciones. El segundo equipo con más apariciones el equipo 8 con 16 apariciones y en tercer lugar el equipo 7 con 12 apariciones. Asimismo, es relevante mencionar que se identificaron variables categóricas para realizar regresiones en Ciencia de Datos, por lo que las columnas “Quarter” y “day” se reemplazaron. Quarter se convirtió en numérica (valores: 1 y 2) y para cada día se creó una nueva columna binaria (0,1).



### Tarea 3 – Exploración de datos (Análisis de Datos)

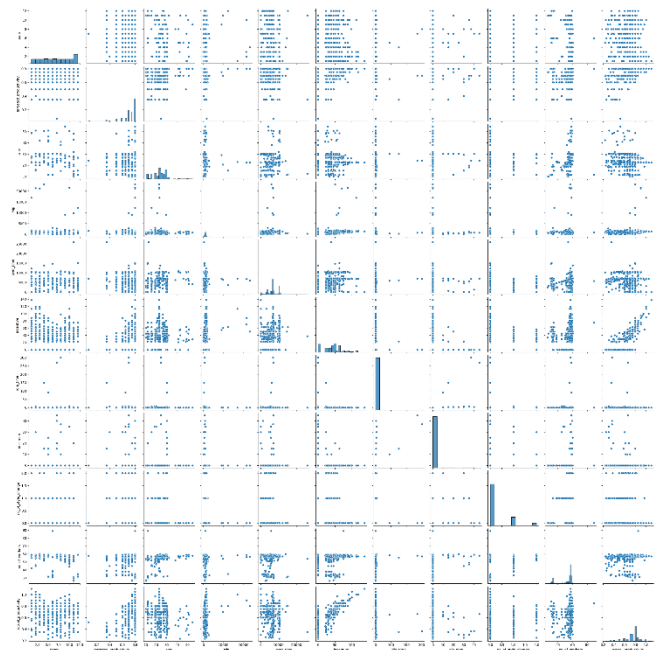
Para el análisis de datos, realizamos diferentes gráficas que se muestran a continuación, que nos permitieron entender mejor el comportamiento de cada variable involucradas en la productividad de manufactura de una empresa. Primero encontramos las **estadísticas descriptivas** para cada variable numérica relacionada con la productividad de diferentes equipos. En la tabla se encuentra lo que obtuvimos. En total encontramos 691 observaciones. La productividad objetivo promedio es de 0.73, indicando que en general se apunta a un alto nivel de eficiencia. El tiempo asignado para una tarea tiene una media de 22.36 y una gran dispersión, lo que sugiere diferencias significativas en la complejidad o el tiempo estándar para realizar las tareas. Por el lado del wip, vemos que tiene una amplia gama de valores, indicando variabilidad en la cantidad de trabajo en proceso. Las horas extras promedian 4569.06, lo que podría reflejar períodos de alta demanda laboral, mientras que los incentivos varían sustancialmente entre los equipos. El tiempo inactivo y el número de trabajadores inactivos son bajos en promedio, lo que podría indicar que se mantiene una buena gestión para evitar la ociosidad. Los cambios de estilo no son frecuentes, con un promedio de 0.62 cambios. El número de trabajadores por equipo varía ampliamente, de 2 a 89, con un promedio de alrededor de 52. Finalmente, la productividad real promedio es de 0.72, cercana a la

	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
count	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000	691.000000
mean	6.520984	0.723980	23.245412	1190.465991	6508.205499	44.483357	1.264834	0.639653	0.260492	52.445731	0.722013
std	3.458333	0.102006	6.976497	1837.455001	2864.505604	27.596591	16.712923	4.283652	0.537185	9.417353	0.154784
min	1.000000	0.070000	10.050000	7.000000	0.000000	0.000000	0.000000	0.000000	0.000000	26.000000	0.233705
25%	4.000000	0.700000	18.790000	774.500000	4560.000000	30.000000	0.000000	0.000000	0.000000	52.000000	0.661470
50%	6.000000	0.750000	22.520000	1039.000000	6840.000000	45.000000	0.000000	0.000000	0.000000	57.000000	0.750608
75%	10.000000	0.800000	28.080000	1252.500000	7200.000000	60.000000	0.000000	0.000000	0.000000	58.000000	0.800402
max	12.000000	0.800000	54.560000	23122.000000	25920.000000	138.000000	300.000000	45.000000	2.000000	89.000000	1.100484

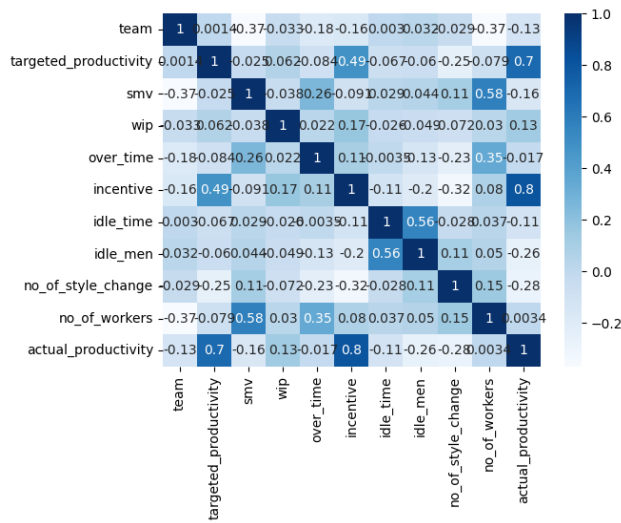
productividad objetivo, pero con una variabilidad que sugiere que algunos equipos superan o no alcanzan las metas establecidas.

Para profundizar mejor estos datos, construimos **histogramas para cada variable y gráficos de dispersión entre todas las variables** como se ve a continuación. En el grafico se puede ver en la diagonal los histogramas de cada una de las variables, mostrando la distribución de cada variable individualmente. Las demás celdas son diagramas de dispersión entre las variables. El histograma de la variable 'team' tiene una distribución relativamente uniforme, indicando que los equipos están distribuidos de manera bastante equitativa en los datos. La variable 'targeted\_productivity' parece tener una concentración alrededor de valores cercanos a 0.7, lo que podría indicar una meta de productividad común. La variable 'idle\_time' tiene una distribución sesgada hacia la izquierda, lo que significa que hay muchos registros con poco o ningún tiempo de inactividad. El 'idle\_men' muestra una distribución similar, lo que sugiere que la mayoría de las observaciones tienen pocos hombres ociosos. Las variables 'incentive' y 'over\_time' también están sesgadas hacia valores más bajos, indicando que hay muchas observaciones con pocos incentivos y tiempos extra bajos, respectivamente. La variable 'no\_of\_style\_change' muestra que la mayoría de los datos tienen cero o uno cambios de estilo, siendo menos frecuentes más cambios. La variable 'incentive' muestra una distribución sesgada hacia valores más bajos, con una cola que se extiende hacia valores más altos. Esto sugiere que, aunque la mayoría de los incentivos son bajos, hay casos con incentivos significativamente mayores. La variable 'wip' (work in progress) presenta una distribución altamente concentrada en valores bajos, con algunos valores extremos muy altos. Las

variables 'targeted\_productivity' y 'actual\_productivity' parecen tener una distribución bimodal, con dos picos distintos, lo que podría indicar la presencia de dos grupos diferenciados dentro de los datos en términos de productividad. En los diagramas de dispersión no se observan patrones claros de correlación lineal fuerte entre las variables, pero sí ciertas tendencias. Por ejemplo, hay una ligera tendencia ascendente entre 'targeted\_productivity' y 'incentive', lo que podría indicar que hay una relación entre una mayor productividad objetivo y mayores incentivos. No obstante, la concentración de puntos en valores bajos de 'idle\_time' y 'idle\_men' dificulta discernir una relación clara con 'targeted\_productivity'. Algo similar ocurre con 'over\_time', donde la mayoría de los puntos están agrupados en valores más bajos, lo que hace difícil identificar una tendencia clara con otras variables. Por otro lado, hay una relación aparentemente positiva entre 'smv' y 'no\_of\_workers', lo que podría indicar que, a mayor valor estándar por minuto, se requiere un mayor número de trabajadores. También se observa una relación lineal positiva entre 'targeted\_productivity' y 'actual\_productivity', lo cual es esperado ya que es probable que la productividad real tienda a seguir la productividad objetivo. Las relaciones entre algunas variables como 'incentive' y 'wip' no parecen mostrar un patrón claro en el diagrama de dispersión, lo que puede sugerir que no hay una relación lineal directa entre estas. Sin embargo, hay una concentración de datos en valores bajos de 'wip', lo que podría dificultar la visualización de una correlación. Para analizar de una mejor forma los gráficos de dispersión, realizamos una **matriz de correlación** presentada a continuación:

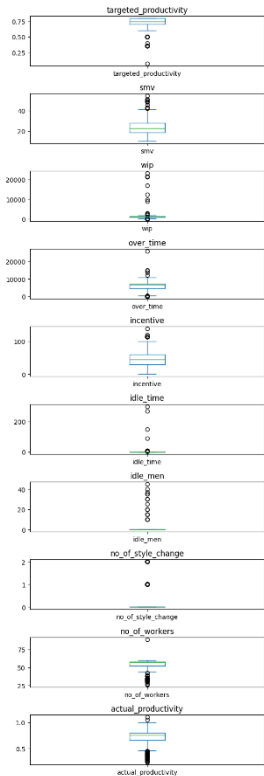


La matriz de correlación presentada revela distintas relaciones entre las variables, destacando una correlación positiva notable entre la productividad actual y la productividad objetivo (0.7), así como con el valor del tiempo asignado para una tarea (0.58) y el número de trabajadores (0.34). Además, se observa una fuerte asociación entre



el tiempo inactivo y la cantidad de operarios ociosos (0.56), lo que sugiere una clara relación entre estos factores. Sin embargo, las horas extras y los incentivos parecen tener una influencia mínima en la productividad real, con correlaciones de -0.017 y 0.08 respectivamente. Curiosamente, la configuración de los equipos no muestra una correlación significativa con otras variables, lo que implica que la composición de los equipos no afecta directamente la productividad en este contexto. Esta información resulta esencial para entender las dinámicas laborales y buscar mejoras en la productividad.

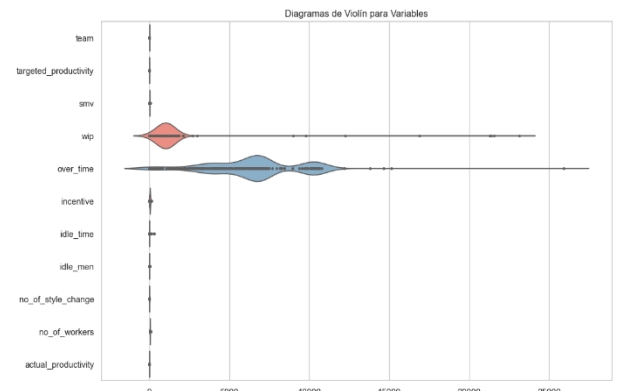
Por otro lado, realizamos un **diagrama de caja** para cada una de las variables, como se muestra en el gráfico del lado. Los diagramas de caja proporcionan una visión clara de la



distribución de los datos para cada una de las variables. En la variable “team”, la caja que está situada relativamente baja y simétrica, indica un numero equilibrado de equipos con una distribución uniforme. La variable productividad esperada muestra una concentración de datos cerca del valor más alto, con algunos valores atípicos por debajo, sugiriendo que la mayoría de los objetivos de productividad están fijados en niveles altos. En el caso del tiempo asignado para una tarea, la distribución es más amplia y simétrica, lo que podría indicar una variación consistente en los valores estándares de tiempo de manufactura. Para el wip y el “over time”, se observan bastantes valores atípicos, lo que podría indicar eventos poco comunes. La variable de incentivos muestra una distribución cercana a los valores más bajos con algunos valores atípicos, lo que refleja que los incentivos no se aplican frecuentemente o son bajos. Por el lado de las variables del tiempo en el que la producción se interrumpió y el número de trabajadores inactivos tienen un número significativo de valores en cero, indicando que hay momentos en los que no hay tiempo muerto o trabajadores inactivos, pero también muestra valores atípicos que podrían ser causa por algo en específico en su momento. En cuanto a la variable de cambios en el estilo de los productos se muestra que la mayoría de las observaciones tienen pocos cambios de estilo, lo cual es deseable en este entorno para mantener la eficiencia. Por el lado de la variable de número de trabajadores por equipo, tiene una variabilidad moderada con valores atípicos que indican equipos de tamaño inusualmente grandes o pequeños. Para finalizar, la variable de productividad actual parece distribuir simétricamente alrededor del valor medio, con valores atípicos a ambos lados, lo que sugiere que, mientras la mayoría de los datos de productividad se concentran alrededor de la media, hay casos de productividad alta o baja.

Por último, realizamos un **diagrama de violín** como se puede ver en el gráfico. En el gráfico se observan dos diagramas de violín sobresalientes para las variables wip y el tiempo extra de cada equipo. Por el lado de la variable wip se ve que sigue una distribución bimodal, indicando dos grupos distintos dentro de los datos. El primer grupo podría representar situaciones en las que el trabajo en curso es bajo o casi nulo, mientras que el segundo grupo podría corresponder a lo contrario. El tener estos dos picos sugiere que existen diferentes condiciones o contextos en los que se encuentra el trabajo en curso. Por otro

lado, la variable del tiempo extra de cada equipo tiene una distribución más amplia y uniforme. Esto indica que las horas extras trabajadas no están agrupadas en valores específicos, sino que se extienden a lo largo de un rango amplio. Puede que algunos trabajadores trabajen



ocasionalmente horas extras, mientras que otros lo hagan con más frecuencia. Por último, las demás variables no muestran una variabilidad significativa en sus datos. Esto se refleja en la falta de un diagrama de violín sobresaliente. Esta diferencia se debe a que estas variables tienen unos valores más consistentes o menos dispersos comparándolos con el wip y el tiempo extra de trabajo.

#### Tarea 4 – Modelamiento (Ciencia de Datos)

Para la ciencia de datos lo primero que se hizo fue eliminar las columnas de las variables categóricas, dejando solo las columnas dummies. Posteriormente se hizo una investigación acerca de las variables que a partir del análisis de los datos parecían tener poco impacto en la productividad, como los incentivos y las horas extras, donde se identificaron dos aspectos importantes. Primero, que los incentivos son una herramienta que tiene un gran impacto en la satisfacción laboral y consecuentemente en la productividad de los trabajadores (Medina, 2012). Segundo, que a pesar del pensamiento común, las horas extras afectan negativamente la productividad, ya que la capacidad de pensamiento y aprendizaje de los trabajadores disminuye (LA REPÚBLICA, 2022). Por estas razones decidimos incluir las variables en el modelo inicial.

Teniendo en cuenta lo anterior, en un principio usamos todas las variables para el modelo inicial, con el fin de poder responder las preguntas ¿Qué factores están afectando la productividad? Y ¿En qué medida? Con esto en mente, separamos los datos, en datos de entrenamiento y en datos de prueba, en una proporción 80%-20% respectivamente. Con los datos de entrenamiento, usamos diferentes métodos de estimación, tales como regresión

```
from sklearn.linear_model import LinearRegression

# crear el objeto del modelo
linreg = LinearRegression()

# ajustar los parámetros del modelo usando los datos de entrenamiento
linreg.fit(X_train, y_train)
```

LinearRegression ⓘ ⓘ  
LinearRegression()

lineal con sklearn, validación cruzada y regresión lineal con statsmodels. Al comparar los tres métodos de predicción, se puede observar que el que menor error cuadrático medio tenía era la regresión con sklearn con un  $MSE = 0.004$ . Sin embargo, se decidió priorizar la regresión de statsmodels, ya que con este es posible evidenciar cuales variables son significativas y cuáles no. El primer modelo, el cual incluye todas las variables, tiene un  $R^2$  ajustado = 0.788 y muestra que el WIP, el número de cambios y el cuarto del mes no son significativas.

Posteriormente, se realizó nuevamente la regresión eliminando las variables que no fueron significativas en el modelo pasado. Con este nuevo modelo, obtuvimos un  $R^2$  ajustado = 0.789 lo cual muestra que las variables escogidas explican en mayor proporción la varianza de la productividad actual y por ende se puede considerar que es un mejor modelo.

OLS Regression Results			
=====			
Dep. Variable:	actual_productivity	R-squared:	0.794
Model:	OLS	Adj. R-squared:	0.788
Method:	Least Squares	F-statistic:	121.0
Date:	Tue, 12 Mar 2024	Prob (F-statistic):	2.13e-160
Time:	14:52:09	Log-Likelihood:	633.82
No. Observations:	518	AIC:	-1234.
Df Residuals:	501	BIC:	-1161.
Df Model:	16		
Covariance Type:	nonrobust		
=====			

Modelo Sweing con todas las variables

OLS Regression Results			
=====			
Dep. Variable:	actual_productivity	R-squared:	0.794
Model:	OLS	Adj. R-squared:	0.789
Method:	Least Squares	F-statistic:	149.3
Date:	Mon, 11 Mar 2024	Prob (F-statistic):	2.48e-163
Time:	19:22:15	Log-Likelihood:	633.09
No. Observations:	518	AIC:	-1238.
Df Residuals:	504	BIC:	-1179.
Df Model:	13		
Covariance Type:	nonrobust		
=====			

Modelo Sweing con variables significativas

Finalmente se realizó el mismo procedimiento descrito anteriormente para el departamento de “Finishing”. Donde se obtuvo un  $R^2$  ajustado = 0.155 para el modelo con todas las variables y un  $R^2$  ajustado = 0.170 para el modelo reducido, por lo que se escogió este segundo modelo, el cual solo cuenta con tres variables significativas, el tiempo extra, el smv y el número de trabajadores. A pesar de que el modelo es significativo, el bajo valor del  $R^2$  ajustado muestra que las variables de entrada explican menos del 20% de la varianza de la productividad



actual, por lo cual sería de interés de la empresa buscar mas factores que puedan estar afectando la productividad de este departamento.

OLS Regression Results			
Dep. Variable:	actual_productivity	R-squared:	0.210
Model:	OLS	Adj. R-squared:	0.155
Method:	Least Squares	F-statistic:	3.826
Date:	Sun, 10 Mar 2024	Prob (F-statistic):	3.51e-05
Time:	20:45:57	Log-Likelihood:	52.393
No. Observations:	186	AIC:	-78.79
Df Residuals:	173	BIC:	-36.85
Df Model:	12		
Covariance Type:	nonrobust		

Modelo finishing todas las variables

OLS Regression Results			
Dep. Variable:	actual_productivity	R-squared:	0.183
Model:	OLS	Adj. R-squared:	0.170
Method:	Least Squares	F-statistic:	13.62
Date:	Sun, 10 Mar 2024	Prob (F-statistic):	4.67e-08
Time:	20:46:19	Log-Likelihood:	49.346
No. Observations:	186	AIC:	-90.69
Df Residuals:	182	BIC:	-77.79
Df Model:	3		
Covariance Type:	nonrobust		

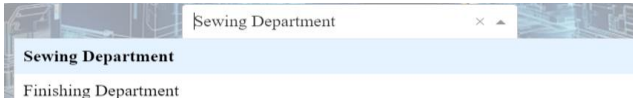
Modelo finishing reducido

### Tarea 5 – Diseño y desarrollo del tablero (Tablero de datos)

Tablero\_De\_Datos (Soporte 4) crea una aplicación web (Dashboard) utilizando la librería Dash en Python. El propósito del código es construir un tablero interactivo que le permita al departamento de producción seleccionar un departamento (Sewing o Finishing) y posteriormente introducir valores para las variables que resultaron significativas del modelo de regresión planteado en Ciencia de Datos y así obtener una predicción para la productividad real del área. El “Layout” de la aplicación se programa utilizando HTML con componentes de CSS para styling. Los “inputs” son:

Un menú desplegable para seleccionar el departamento de costura o acabado:

El departamento de producción podrá calcular la potencial productividad real de alguno de los dos departamentos a partir de variables características del proceso. Para esto hay boxes con entradas para las variables que fueron caracterizadas como estadísticamente significativas en la tarea de Ciencia de Datos. Se observan los boxes para incluir los inputs de las variables y el botón de “calcular” para realizar el cálculo correspondiente.



Sewing Department

team

targeted\_productivity

smv

over\_time

incentive

idle\_time

idle\_men

no\_of\_workers

day\_Sunday

day\_Saturday

day\_Monday

day\_Tuesday

day\_Wednesday

day\_Thursday

Calcular

Finishing Department

smv

over\_time

no\_of\_workers

Calcular

Se utiliza CSS dentro del componente html.Div de forma inline (atributo style) para poner una imagen opaca de produccion, poner el logo de la Universidad de Los Andes en el cuadrante superior izquierdo, cambiar la letra y la fuente. Para esta aplicación se realizan tres **callbacks**, el **primer callback** es para que se seleccione correctamente el departamento a evaluar y los input boxes sean los correspondientes al departamento. Esto se logra por medio de una función que se creó, llamada `update_x_values_input(department)`. Esta tiene como parámetro el departamento seleccionado en la lista desplegable y retorna los inputs deseados.

El **segundo callback** es para que el modelo de regresión haga la predicción de la productividad a partir de los inputs establecidos para cada departamento. Como se puede observar en la imagen se calcula la productividad y por medio del callback cada vez que se inserten valores y se haga click en submit se actualicen los resultados:

Sewing Department

11

0.97

12

960

78

0

2

12

0

0

0

1

0

0

Calcular

Productividad estimada: 0.9581098548130963

Finishing Department

11

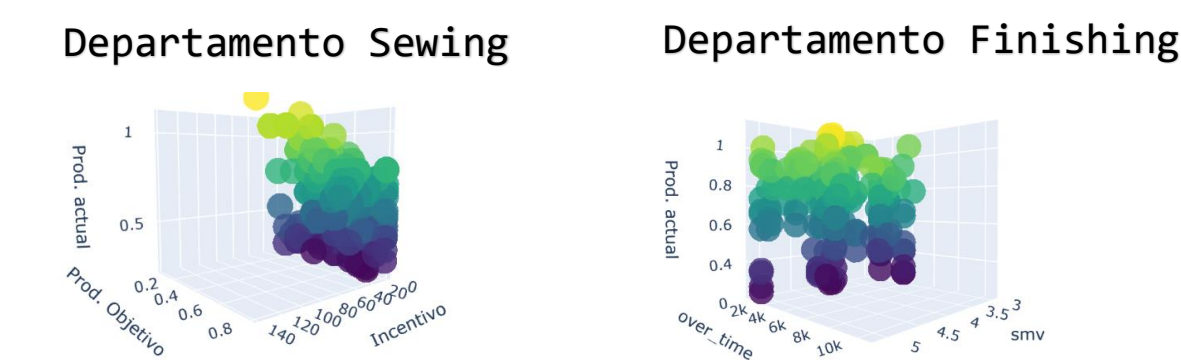
50

3

Calcular

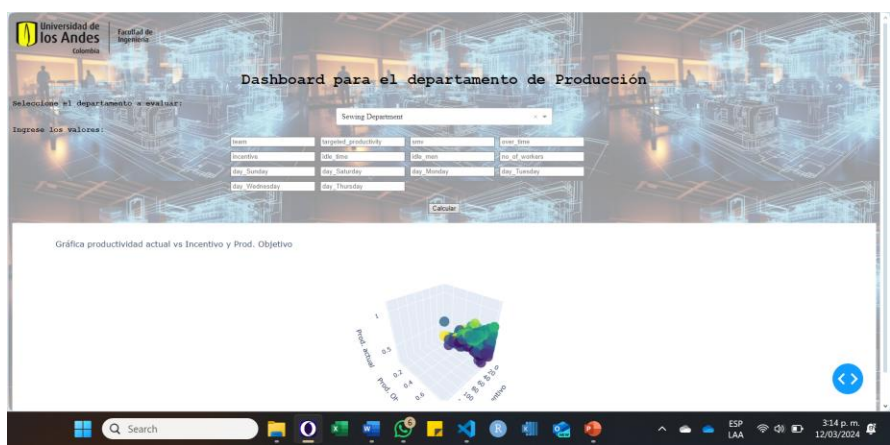
Productividad estimada: 0.8716749807293643

El **tercer callback** actualizará la gráfica 3D de la parte inferior. De acuerdo con el departamento seleccionado se genera una gráfica 3D para visualizar el comportamiento de algunas de las variables alrededor de la variable de respuesta. Para incluir las gráficas en tres dimensiones en el app *layout* se incluye usando *html.Div* el argumento *dcc.graph* con el id de la gráfica que genera el output de la función `update_3d_plot(department)`. Se observa en las imágenes



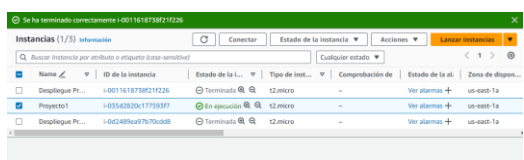
La gráfica le permite al usuario seleccionar las burbujas y ver cuál es el comportamiento de la variable de respuesta y, cuando las variables explicativas x toman un valor específico.

El Dashboard completo se ve de la siguiente manera:

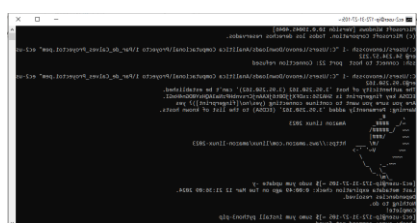


## Tarea 6 – Despliegue (Despliegue)

Para el despliegue iniciamos lanzando una instancia usando el servicio EC2 de AWS. Para esto creamos la siguiente instancia, que tiene una dirección IP(v4) publica 3.95.250.162. A continuación se puede ver la instancia en AWS:



Una vez esta estaba en ejecución, desde la terminal la vinculamos con la instancia EC2 y logramos entrar a la maquina virtual. A continuación, se ve la terminal:



Estando ahí, realizamos la instalación de todas las librerías que requeríamos para que corriera el dash y adicional a esto, subimos los archivos a la maquina en la nube. Dentro de los documentos que subimos fueron “Tablero\_De\_Datos.py”, “datosproyectos.csv” y “finishinfdata.csv”. Por último, realizamos una pequeña modificación sobre el archivo “Tablero\_De\_Datos.py” donde le agregamos lo siguiente en una línea de código: “`app.run_resrver(host=”0.0.0.0”, debug=True)`”. Teniendo esto último, agregamos la regla de entrar a nuestro grupo de seguridad en la red y con esto lanzamos el tablero en el servidor. El URL es el siguiente: `http://3.95.250.162:8050/`