



Análisis de la evolución del mercado laboral en EE. UU. entre 1995 y 2011, su relación con las características demográficas de la población y su impacto en los servicios de apoyo laboral

Integrantes:

José Luis Santamaria Andrade - 22502265

Nicolas Lozano Mazuera - 22500565

Martín Alonso Herrera Vargas - 22501540

Luisa María Candelo Angulo – 22500699

Grupo 2

Curso de ETL y Ciencia de Datos - 551464

Fernando Barraza Alvarado

Abril, 2025

Análisis de la evolución del mercado laboral en EE. UU. entre 1995 y 2011, su relación con las características demográficas de la población y su impacto en los servicios de apoyo laboral

1. Descripción del problema

Con una población de aproximadamente 340 millones de personas, Estados Unidos es el tercer país más poblado del mundo después de China e India (Banco Mundial, 2023). Esta magnitud demográfica, sumada a su rol como la mayor potencia económica global, plantea desafíos significativos en el mercado laboral, especialmente en la integración de trabajadores desempleados y la provisión de servicios de apoyo eficaces. En 2023, su PIB nominal alcanzó los 27.7 billones de dólares, representando aproximadamente el 25% del PIB mundial, con un PIB per cápita de 82,769 dólares, lo que lo posiciona como una de las naciones más prósperas (International Monetary Fund, 2023). Estas condiciones económicas, junto con la estabilidad del país, han convertido a Estados Unidos en un destino altamente atractivo para la migración. Actualmente, alrededor del 15% de su población total (50 millones de personas) es de origen extranjero, de los cuales el 77% reside legalmente (ONU Migración, 2024). Además, el país lidera el envío de remesas a nivel mundial, lo que refleja la importancia de su mercado laboral en la movilidad económica global. En este contexto, comprender cómo las dinámicas demográficas y los servicios de apoyo al empleo influyen en las tasas de desempleo y en la duración de la desocupación es fundamental para diseñar políticas públicas más efectivas que promuevan la estabilidad y el crecimiento económico.

2. Definición del problema

En este proyecto se analizarán las características demográficas de la población participante del programa de empleabilidad de Wagner Peyser en EE. UU. entre 1995 y 2011 y el impacto de este servicio de apoyo en el mercado laboral.

3. Justificación del proyecto

Este proyecto tiene como objetivo analizar la evolución de la empleabilidad en Estados Unidos entre 1995 y 2011, considerando factores clave como el volumen de la población, la edad promedio de la fuerza laboral, el género y el nivel de escolaridad, con una segmentación trimestral. A través de este estudio, se podrá evaluar la evolución de la tasa de desempleo tanto a nivel nacional como estatal, identificar tendencias en el envejecimiento de la población trabajadora y examinar los cambios en la proporción de niveles de escolaridad dentro de la fuerza laboral. Además, el análisis permitirá observar la relación entre el desempleo y los servicios de apoyo gubernamentales. Estos hallazgos no solo serán relevantes para la población estadounidense, sino también para ciudadanos de países como Colombia, dado que Estados Unidos es el principal destino de los migrantes colombianos, concentrando el 34.6% del total de emigrantes. La información obtenida contribuirá al diseño de políticas públicas más eficaces para mejorar los programas de empleo y apoyo laboral, reduciendo las tasas de desempleo y fortaleciendo la protección

de los trabajadores desempleados. Asimismo, proporcionará un panorama más claro sobre las expectativas económicas que pueden tener los migrantes en función de su perfil sociodemográfico (género, edad, nivel de escolaridad) y del estado de destino dentro de EE.UU.

4. ETL

A continuación, se describe el proceso de exploración inicial de datos, extracción, transformación y validación aplicado en el proyecto de análisis de las características demográficas de la población participante del programa de empleabilidad de Wagner Peyser en EE. UU. entre 1995 y 2011. El objetivo del proceso ETL es consolidar los datos provenientes de diversas fuentes, limpiarlos y organizarlos para su posterior análisis.

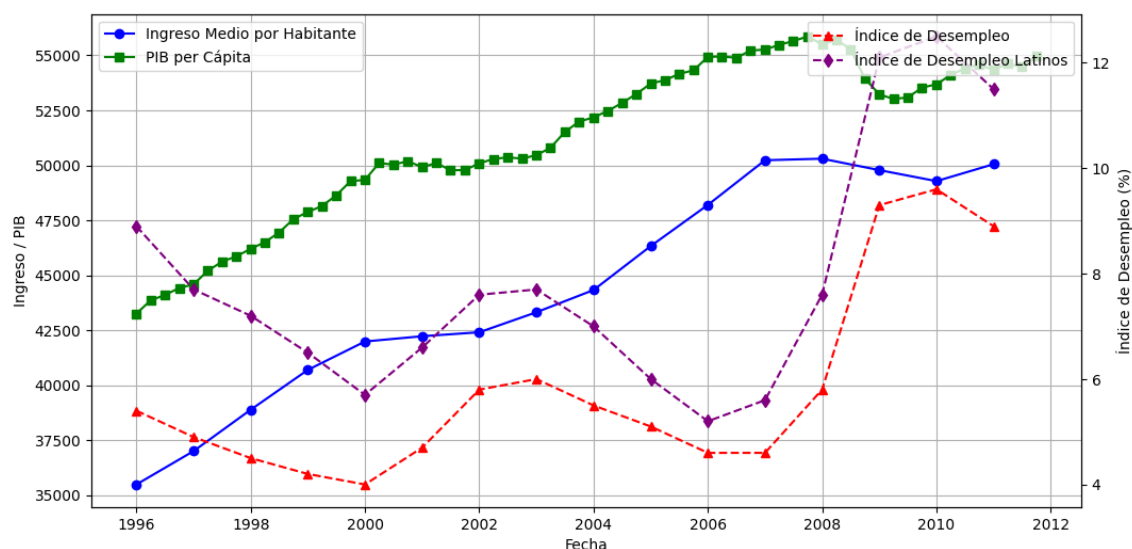
4.1. Exploración inicial de datos

Fuente Principal:

[Employment Research Data Center | Upjohn Institute](#): Esta fuente muestra la información general de la fuerza laboral de cada estado de Estados Unidos durante los años de 1995 y 2011 y las características demográficas de las personas que se postularon al programa Wagner Peyser durante ese mismo período de tiempo.

Fuentes complementarias de información:

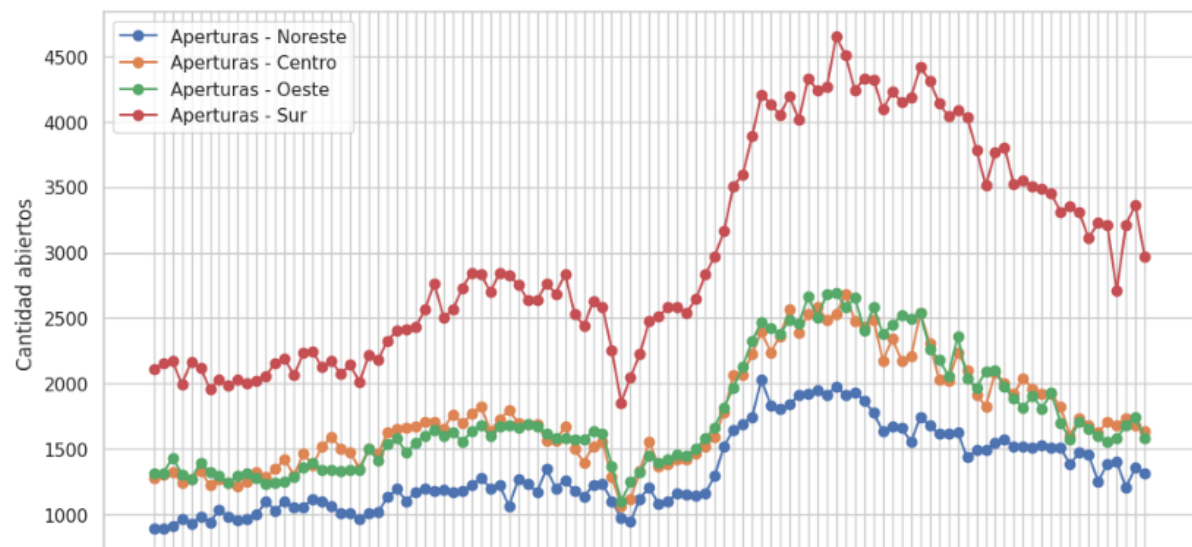
[Datos económicos de la Reserva Federal | FRED | Fed de San Luis](#): En la fuente de información del Banco de la Reserva Federal de St. Louis se extrajo información nacional de cuatro variables: promedio de ingresos anuales por habitante, el PIB per cápita, índice de desempleo total e índice de desempleo para la población hispana. Esta información se cargó al entorno colaborativo y se convirtió a un dataframe, donde se nombraron las columnas de una forma fácil de interpretar, se validó la tipología de los datos haciendo las transformaciones necesarias y posteriormente se realizó un gráfico para una mejor interpretación de los datos:



Gráfica 1. Evolución del ingreso medio por habitante, PIB per cápita e índice de desempleo

Desde el año de 1996 el ingreso medio por habitante y el PIB per cápita incrementan de una manera casi constante, salvo por los años 2009 y 2010 en donde experimenta una leve caída. Entre los años de 2007 y 2009 hay un fuerte impacto en el índice de desempleo para la población en general y en una mayor proporción para la población hispana. Lo que muestra la importancia de tener claridad de la situación laboral en Estados Unidos y del impacto de los servicios de apoyo laboral.

U.S. Bureau of Labor Statistics : La fuente de información utilizada es un documento en formato XLSX proporcionado por la **Bureau of Labor Statistics (BLS)**. Este archivo contiene múltiples hojas que detallan la cantidad de ofertas laborales, contrataciones, procesos iniciados y despidos, organizados mes a mes desde el año 2016 hasta 2024. La información está segmentada tanto por actividad económica, como la construcción o la agricultura, como por regiones geográficas. Para el análisis, se priorizó la información desglosada por región, considerando las principales zonas del país: noreste, centro, sur y oeste. En estas regiones, se analizaron variables clave como la cantidad de ofertas laborales, el número de contrataciones y la cantidad de despidos a lo largo del período mencionado.



Gráfica 2. Evolución de aperturas por región

4.2. Extracción

Los datos utilizados en este proyecto provienen del conjunto "Public Workforce System Dataset", la cual contiene información detallada sobre el empleo en 51 estados de los EE.UU., abarcando el período desde el primer trimestre de 1995 hasta el cuarto trimestre de 2011. Con un total de 3.468 observaciones, la base incluye 28 variables relacionadas con la fuerza laboral, el empleo, el desempleo, y los servicios de apoyo a los desempleados. Se construye la variable 29 de nombre "Mujeres candidatas al programa Wagner Peyser", con el objetivo de analizar la proporción de hombre y mujeres que se postulan al programa. Los datos fueron obtenidos en formato CSV desde la plataforma oficial. Se empleó la librería pandas de Python para leer estos archivos con los delimitadores adecuados y almacenarlos en estructuras de datos tipo DataFrame. Además, se realizó una inspección inicial de los archivos para identificar inconsistencias.

Relación de variables en el data set			
Número columna	Nombre variable	Non-Null Count	Tipo de dato
1	Estado	3468	object
2	Año	3468	int64
3	Trimestre	3468	int64
4	Total fuerza laboral	3468	int64
5	Personas empleadas	3468	int64
6	Personas desempleadas	3468	int64
7	Índice de empleabilidad	3468	object
8	GDP Trimestre	2244	float64
9	Beneficios pagados	3468	int64
10	Reclamaciones iniciales	3468	int64
11	Pagos iniciales	3468	int64
12	Semanas reclamadas	3468	int64
13	Semanas compensadas	3468	int64
14	Total candidatos al programa Wagner Peyser	3380	float64
15	Hombres candidatos al programa Wagner Peyser	3381	float64
16	Total candidatos elegibles para el programa Wagner Peyser	3380	float64
17	Personas empleadas que aplicaron al programa	3361	float64
18	Personas desempleadas que aplicaron al programa	3380	float64
19	Candidatos con nivel de escolaridad cursando High School	3335	float64
20	Candidatos con nivel de escolaridad menor a High School	3380	float64
21	Candidatos con nivel de escolaridad High School	3380	float64
22	Candidatos con nivel de escolaridad Postsecondary	3354	float64
23	Candidatos menores de 19 años	3379	float64
24	Candidatos entre 19 y 22 años	3381	float64
25	Candidatos entre 22 y 44 años	3381	float64
26	Candidatos entre 45 y 54 años	3381	float64
27	Candidatos mayores de 55 años	3381	float64
28	Total personas empleadas por el programa	3192	float64
29	Mujeres candidatas al programa Wagner Peyser	3380	float64

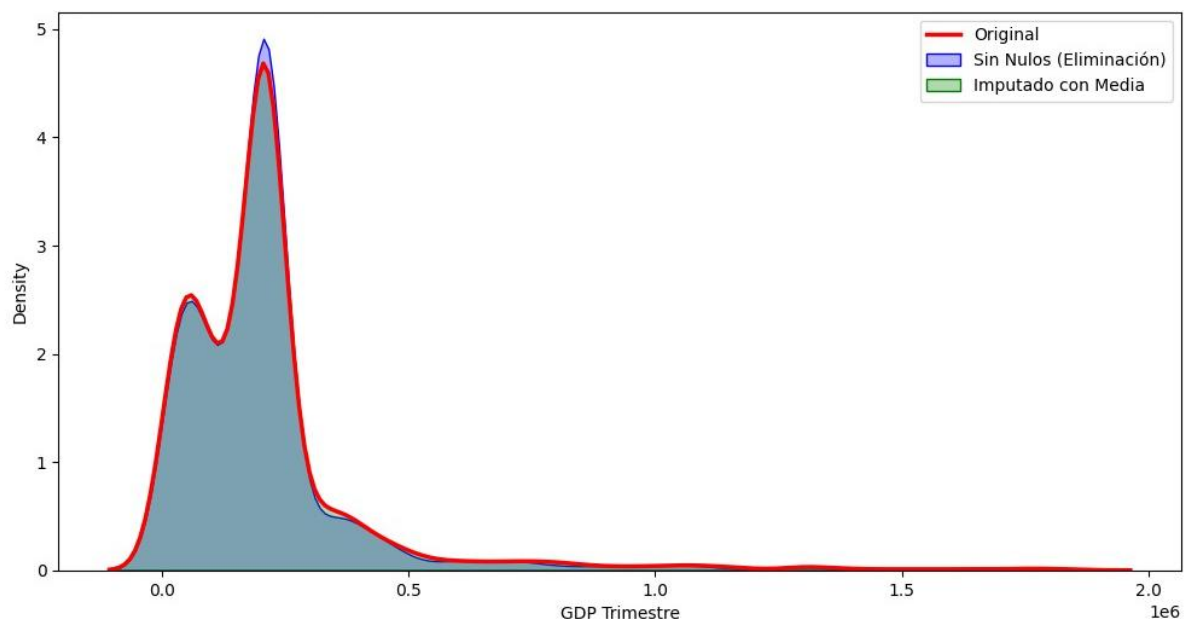
Tabla 1. Variables observadas

4.3. Transformación

En la fase de transformación, se realizaron procesos de limpieza, modificación y detección de anomalías con el objetivo de mejorar la calidad, consistencia y utilidad de los datos extraídos. Inicialmente, se realizó una exploración de los datos, en donde se vio la necesidad de estandarizar los encabezados de las columnas para facilitar su interpretación y manipulación. Los nombres estaban en inglés, por lo que fueron traducidos y

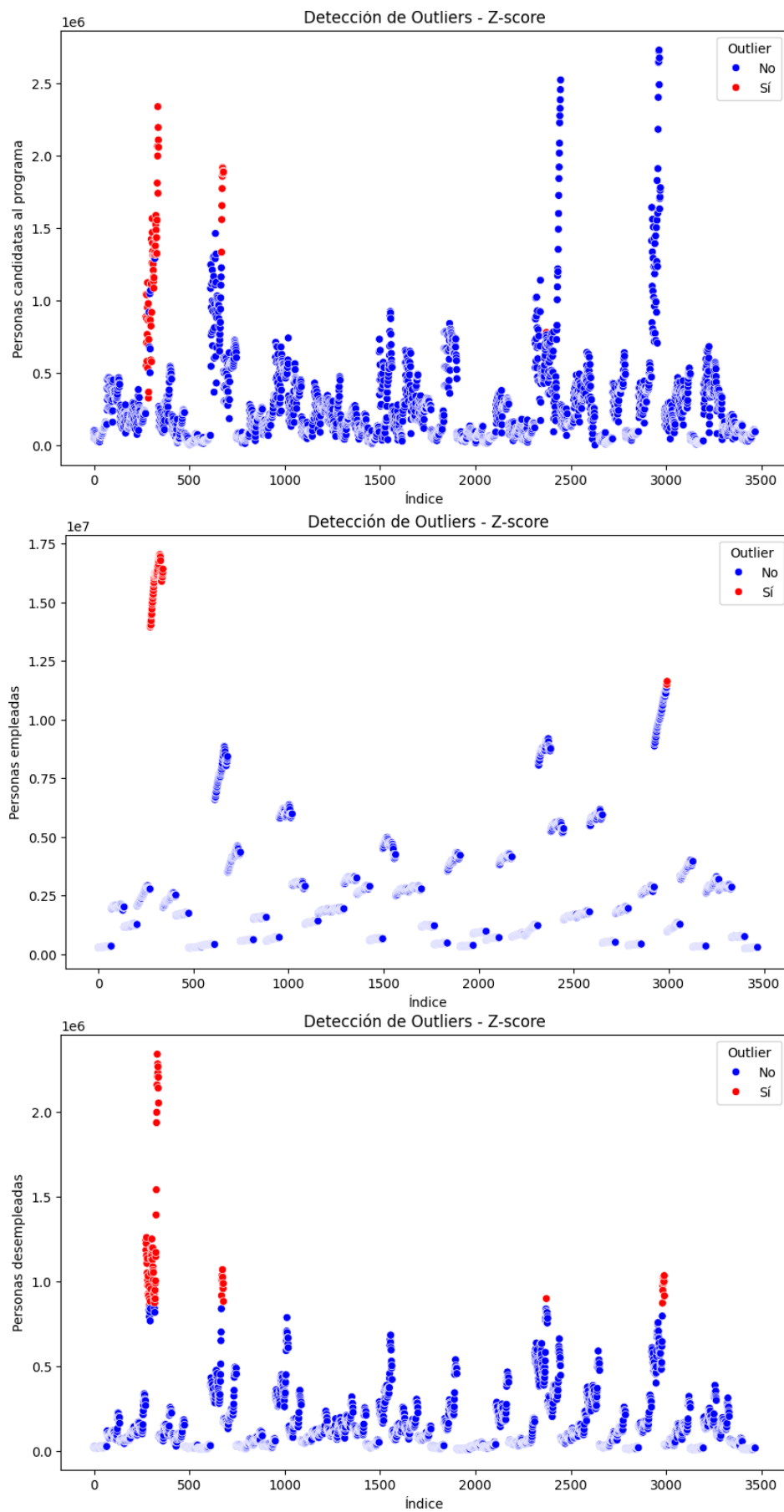
homogeneizados. Además, se creó una nueva columna denominada "Candidatas mujeres" para mejorar la segmentación de los datos, dado que solo se disponía de una columna con el total de candidatos y otra con el total de candidatos hombres. El cálculo de esta columna se realizó restando el total de candidatos menos los hombres candidatos.

Posteriormente, se identificaron y modificaron los tipos de datos de algunas variables para garantizar su coherencia y prevenir errores en el análisis posterior. Se detectaron variables con valores en formato decimal que debían ser convertidas a enteros para evitar imprecisiones. Esta conversión se aplicó a columnas relacionadas con el total de fuerza laboral, personas empleadas y desempleadas, así como a diversas categorías del programa Wagner Peyser. Adicionalmente, se identificaron valores nulos en el conjunto de datos y se implementó la imputación con la media para el caso de la variable PIB trimestral.



Gráfica 3. Impacto de los valores nulos en la distribución del PIB trimestral

Finalmente, como parte del proceso de transformación, se realizó la detección de valores atípicos (outliers) utilizando la técnica del Z-score. Esta metodología permite identificar valores cuya desviación estándar supera un umbral predeterminado, generalmente establecido en 3 sigmas. Los outliers corresponden a observaciones que se desvían significativamente de la media en términos de desviaciones estándar, lo que puede deberse a errores de medición, diferencias en la escala de los datos o la presencia de valores extremos reales. Sin embargo, dado que la fuente de información utilizada proviene de datos gubernamentales oficiales sobre el mercado laboral de EE. UU., se determinó que los valores identificados como outliers no eran anomalías generadas por errores, sino que representaban eventos reales dentro del período de análisis. Por esta razón, no se aplicó ninguna técnica de tratamiento y los valores fueron conservados en el conjunto de datos.



Gráfica 4. Detección de Outliers – Z-score

4.4. Validación

Tras completar la fase de transformación, se llevaron a cabo distintas verificaciones para asegurar la calidad e integridad de los datos. En primer lugar, se revisó la coherencia interna de la información, comprobando que la suma de personas empleadas y desempleadas coincidiera con el total de la fuerza laboral en cada registro, y que los diferentes grupos etarios (menores de 19 años, entre 19 y 22, entre 22 y 44, entre 45 y 54 y mayores de 55) reflejaran correctamente el total de candidatos al programa Wagner Peyser, admitiendo pequeñas diferencias (entre -2 y 2) atribuibles al redondeo.

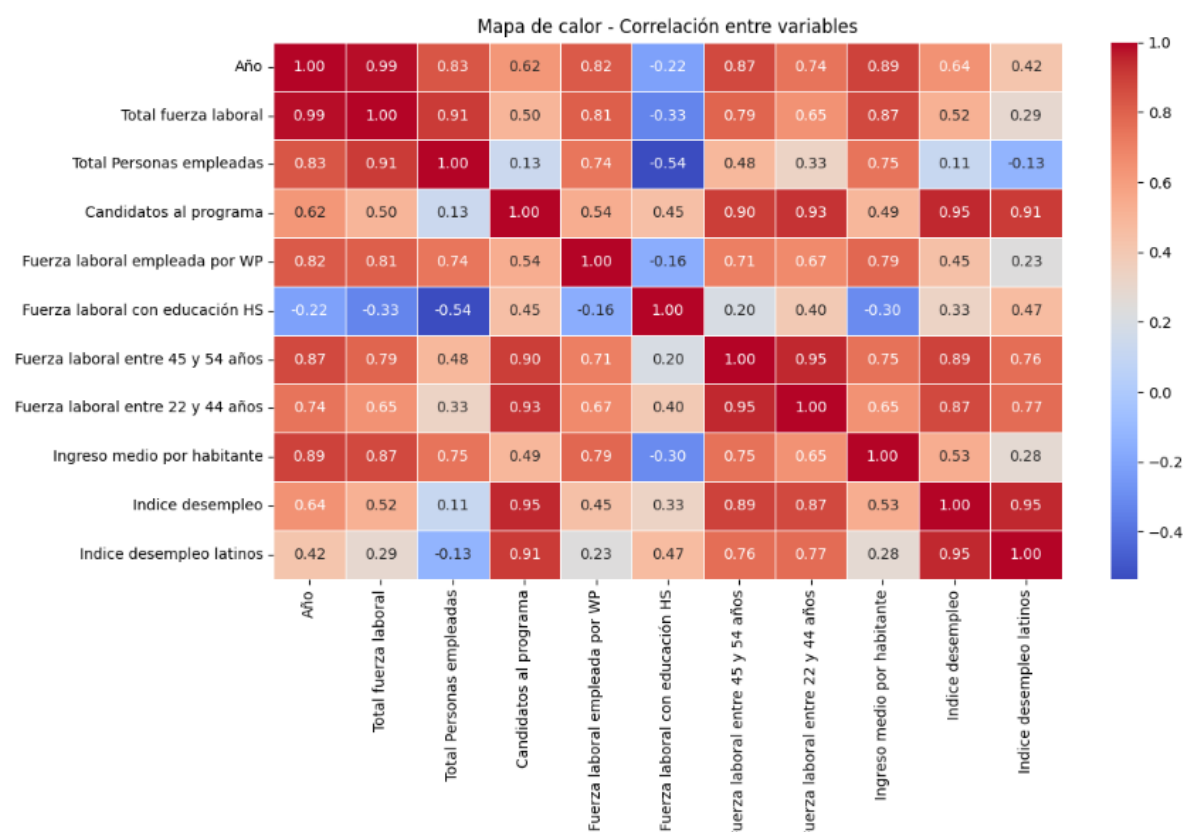
En cuanto al control de valores nulos, tras la imputación de la media en la variable del PIB trimestral, se verificó que no quedaran datos ausentes y se comparó la distribución resultante para descartar la presencia de sesgos significativos. Finalmente, se realizó una inspección visual y estadística de la información, mediante el cálculo de medidas descriptivas (media, mediana, desviación estándar) y la generación de gráficos, con el objetivo de identificar outliers o inconsistencias remanentes. Con estas acciones se confirmó que los datos finales reunían las condiciones necesarias para un análisis confiable.

5. Integración de fuentes de datos

Como parte del proceso de integración de datos dentro del pipeline ETL, se realizó la consolidación de información proveniente de múltiples fuentes, incluyendo tanto la base principal del programa Wagner Peyser como datos económicos del Banco de la Reserva Federal de St. Louis (FRED). En una primera etapa, se aplicaron técnicas de transformación y agregación a la base principal para obtener, por cada año entre 1995 y 2011, variables clave como el total de fuerza laboral nacional, personas empleadas, candidatos con nivel de educación secundaria (High School), distribución por grupos de edad (22-44 años y 45-54 años), total de personas que aplicaron al programa y aquellas que obtuvieron empleo a través del mismo. Simultáneamente, se cargaron cuatro conjuntos de datos adicionales desde FRED, correspondientes al ingreso medio por habitante, el índice de desempleo nacional, el índice de desempleo de migrantes latinos y el PIB per cápita. Estas bases de datos fueron transformadas mediante la estandarización del formato de fecha, extrayendo el componente "Año" para facilitar su alineación con la base principal. Una vez armonizados los formatos, los cuatro datasets de la FRED fueron fusionados en un único DataFrame. Posteriormente, se realizó la integración final mediante uniones externas entre el resumen anual de la base Wagner Peyser y el conjunto económico agregado, lo cual permitió construir un DataFrame consolidado que recoge, de forma integrada, variables del mercado laboral y macroeconómicas por año. Finalmente, se implementó una estrategia de imputación basada en la media estadística para manejar los valores faltantes y garantizar la completitud de la información para los análisis posteriores, con el objetivo de encontrar nuevos hallazgos sobre la relación de las características demográficas de la población, el índice de desempleo para el total de la población y la población latina; y el impacto del programa Wagner Peyser en estas variables.

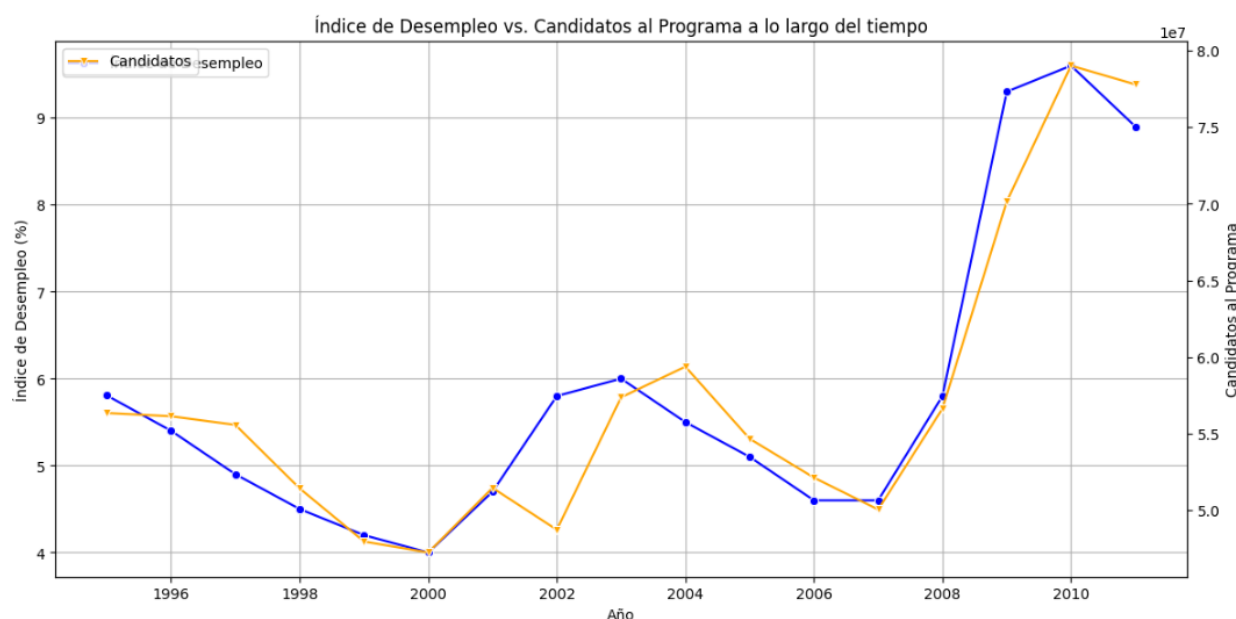
6. Visualización de resultados del Pipeline

Con la integración de los datos del programa Wagner Peyser y los indicadores macroeconómicos provenientes de la Reserva Federal de EE. UU. (FRED), se realizó una exploración de las relaciones clave entre la coyuntura económica y el comportamiento del mercado laboral.



Gráfica 5. Mapa de calor – correlación entre variables

Con el objetivo de analizar la fuerza y dirección de las relaciones lineales entre las distintas variables del dataset resultante del proceso ETL, se realizó un mapa de calor que representa la correlación entre las variables. En la Gráfica 5 se puede observar que existe una correlación positiva (0.95) entre el índice de desempleo general y el número de candidatos al programa Wagner-Peyser. Esto demuestra que, en años con mayor desempleo, más personas buscaron apoyo en este servicio de empleabilidad. Asimismo, la correlación entre los candidatos al programa y el índice de desempleo de la población latina (0.91) también es bastante fuerte, lo que sugiere que esta población ha sido altamente dependiente de los servicios del programa durante años de mayor vulnerabilidad laboral. Finalmente, resulta interesante observar la correlación negativa entre la proporción de fuerza laboral con educación secundaria (high school) y el total de personas empleadas (-0.54), ya que esto puede indicar que los trabajadores con menor nivel educativo enfrentan mayores barreras de acceso al empleo formal, resaltando la necesidad de programas de formación complementaria.

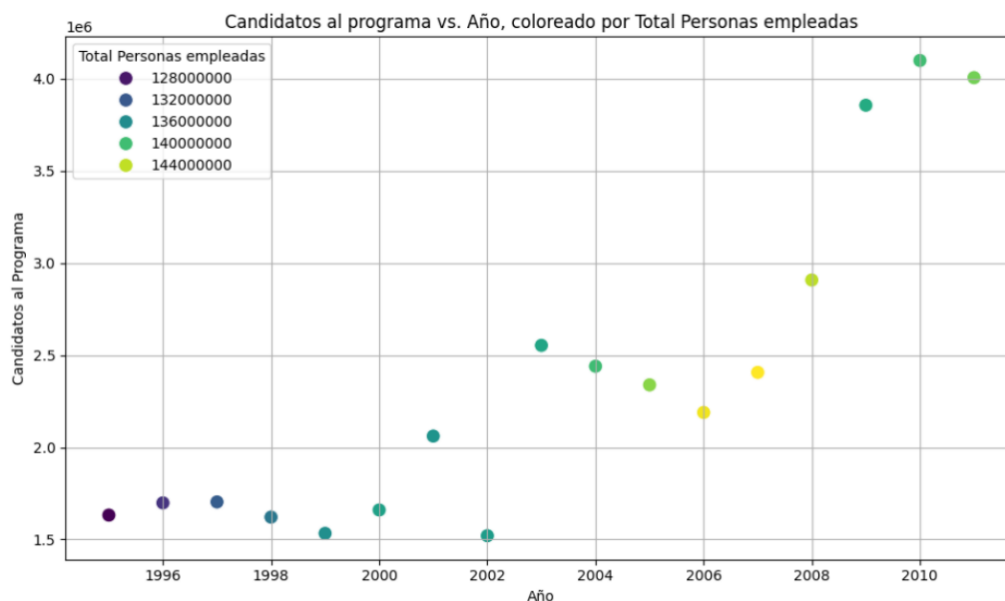


Gráfica 6. Índice de desempleo vs. Candidatos al Programa

En la Gráfica 6 se puede observar la evolución del índice de desempleo general en EE. UU. y la cantidad de personas que se postularon al programa de empleabilidad Wagner-Peyser desde 1995 hasta 2011. En el periodo de análisis se puede apreciar que hay una relación directa entre ambas variables, especialmente en periodos de crisis económica como el año 2008, momento en el cual el desempleo aumentó abruptamente debido a la crisis financiera global. Este fenómeno también se reflejó en el notable incremento en el número de candidatos al programa, lo cual evidencia la función contracíclica del mismo.

A lo largo del periodo analizado, puede observarse que, tras cada incremento en el desempleo, la cantidad de candidatos al programa tiende a subir, lo que sugiere que los ciudadanos perciben esta política como una herramienta confiable en momentos de incertidumbre. Posteriormente, a medida que el desempleo disminuye, la cantidad de candidatos también se reduce, lo cual refuerza su carácter de instrumento de respuesta a momentos de crisis del mercado laboral.

El programa Wagner-Peyser fue establecido mediante la Ley Wagner-Peyser de 1933, la cual forma parte de una serie de políticas diseñadas para aliviar los efectos de la Gran Depresión en los Estados Unidos. Su propósito principal es facilitar la intermediación laboral entre empleadores y desempleados a través de servicios gratuitos de asesoría, orientación profesional, preparación laboral y conexión con oportunidades de trabajo (U.S Department of Labor, 2023). Este programa opera en conjunto con los centros de empleo locales y estatales, conocidos como American Job Centers, asegurando un acceso amplio y equitativo a sus servicios. Además, tiene como objetivo fortalecer el mercado laboral mediante la rápida canalización de trabajadores hacia posiciones disponibles, mejorando la eficiencia y flexibilidad económica. Este servicio es particularmente importante en contextos de desempleo elevado, ya que actúa como un amortiguador ante choques económicos al canalizar rápidamente a los trabajadores afectados hacia nuevas oportunidades laborales.



Gráfica 7. Total de Personas Empleadas vs. Candidatos al Programa, año tras año

En la Gráfica 7 se visualiza la progresión del número de candidatos al programa a medida que avanza el tiempo (eje X), coloreando cada punto según el total de personas empleadas (leyenda a la izquierda). Esta representación multivariable confirma la misma correlación observada en la gráfica previa (entre el desempleo y la cantidad de candidatos), puesto que, en ambos casos, períodos con menor estabilidad laboral coinciden con un aumento en la demanda del servicio. Es importante resaltar el descenso notable que se aprecia en el año 2002: no existe un factor único que explique esa disminución, pero, según la documentación oficial (GAO, DOL), muchos estados comenzaron a integrar programas como Wagner-Peyser y WIA en centros unificados (One-Stop Career Centers) y registraban a las personas bajo un sistema conjunto. Este cambio de metodología, junto con otros factores económicos (como el ajuste tras la burbuja puntocom y los efectos posteriores al 9/11), puede haber reducido la cifra reportada específicamente para “Wagner-Peyser,” aun cuando el total de ciudadanos atendidos por los centros de empleo no disminuyera realmente.

7. Reflexiones finales y conclusiones

El desarrollo del presente pipeline ETL permitió integrar y transformar datos heterogéneos provenientes de diversas fuentes, incluyendo datos del programa Wagner-Peyser y del Banco de la Reserva Federal (FRED), logrando consolidar un dataset robusto y limpio, apto para análisis exploratorios y visualizaciones significativas. Este proceso implicó tareas clave como extracción, transformación, limpieza, imputación de valores nulos e integración de múltiples fuentes de datos.

Desde una perspectiva analítica, los resultados muestran que el programa Wagner-Peyser cumple un papel fundamental en la mitigación del desempleo, al responder eficientemente

frente a las crisis económicas. La alta correlación entre el desempleo y la cantidad de personas que acceden al programa refuerza su importancia como mecanismo institucional de apoyo laboral. Además, la correlación observada con variables como el desempleo latino y el nivel educativo aporta evidencia para el diseño de políticas más focalizadas, que reconozcan los retos específicos de cada grupo poblacional en el mercado de trabajo.

Finalmente, este ejercicio demostró la importancia de los procesos ETL en proyectos de análisis de datos. Una integración efectiva de fuentes y una transformación adecuada de los datos no solo mejora la calidad del análisis, sino que también potencia la toma de decisiones basada en evidencia.

8. Referencias

- Banco Mundial. (2023). *Grupo Banco Mundial*. Obtenido de <https://datos.bancomundial.org/pais/estados-unidos>
- Federal Reserve Bank of St. Louis. (2024). *Federal Reserve Economic Data*. Obtenido de <https://fred.stlouisfed.org/>
- Federal Reserve. (s.f.). *Economic Research Data*. Obtenido de <https://www.federalreserve.gov/econres/economic-research-data.htm>
- International Monetary Fund. (2023). *IMF Data Portal*. Obtenido de <https://www.imf.org/en/Data>
- ONU Migración. (2013). *Perfil Migratorio de Colombia*.
- ONU Migración. (2024). *Informe sobre las migraciones en el mundo 2024*.
- Organización para la Cooperación y el Desarrollo Económico (OCDE). (2024). *Migración*.
- U.S Department of Labor. (2023). *Wagner-Peyser Program*. Obtenido de <https://www.dol.gov/agencies/eta/american-job-centers/wagner-peyser-program>
- U.S. Government Accountability Office (GAO). (2001). *Workforce Investment Act: Better Guidance Needed to Address Concerns Over New Requirements*.
- United States Department of Labor. (s.f.). *U.S. Bureau of Labor Statistics*. Obtenido de <https://www.bls.gov/bls/newsrels.htm#OEUS>
- United States government. (s.f.). *Census Bureau's United States* . Obtenido de <https://www.census.gov/>
- W.E. Upjohn Institute for Employment Research. (s.f.). *Employment Research Data Center*. Obtenido de <https://www.upjohn.org/data-tools/employment-research-data-center>