



**INSTITUTO SUPERIOR
TECNOLÓGICO QUITO**

Formamos tu PROPÓSITO DE VIDA

**YO
SOY
ITQ**



PROYECTO FINAL DE ASIGNATURA

Desarrollo de Software

WWW.ITQ.EDU.EC



**INSTITUTO SUPERIOR
TECNOLÓGICO QUITO**
Formamos tu PROPÓSITO DE VIDA

PROYECTO FINAL DE ASIGNATURA

“Estadística Descriptiva”

“PAOLA ALEXANDRA OLMEDO VILCA”

“Germán del Río y Kevin Jaramillo”

Septiembre 2024 – Febrero 2024





CONTENIDO

OBJETIVO DE LA ASIGNATURA	5
INDICACIONES GENERALES	5
I INTRODUCCIÓN.....	6
1. CAPÍTULO I	7
1.1. NOMBRE DEL PROYECTO	7
1.2. PLANTEAMIENTO DEL PROBLEMA.....	7
1.3. OBJETIVOS GENERALES	8
1.4. OBJETIVOS ESPECÍFICOS.....	8
1.5. JUSTIFICACIÓN	8
2. CAPÍTULO II	10
2.1. MARCO TEÓRICO	10
3. CAPÍTULO III	13
3.1. DESARROLLO Y/O IMPLEMENTACIÓN	13
3.1.1. Metodología	13
3.1.2. Fases de Implementación.....	14
3.1.3. Ajustes y Modificaciones.....	15
3.1.4. Resultados Preliminares.....	15
4. CAPÍTULO IV	16
4.1. ANÁLISIS DE RESULTADOS	16
4.1.1. Presentación de Resultados.....	16
4.1.2. Comparación con Objetivos.....	17
4.1.3. Discusión.....	17
4.1.4. Limitaciones.....	18
4.1.5. Implicaciones	18
5. CAPÍTULO V	19
5.1. CONCLUSIONES.....	19



5.2. RECOMENDACIONES.....	20
6. BIBLIOGRAFÍA.....	23



OBJETIVO DE LA ASIGNATURA

“Analizar la distribución de salarios por sector económico:

Realizar la comparación de la distribución de salarios en diferentes sectores, utilizando medidas de tendencia central, dispersión y gráficos como histogramas y diagramas de Pareto para visualizar los datos.”

INDICACIONES GENERALES

- Respetar el formato establecido por el docente.
- Las imágenes tienen que ir con descripciones y fuentes, con normas APA 7ma edición.
- El presente documento tiene por defecto configurado los diferentes recursos de normas APA, en la sección de estilos, usarlos para los diferentes, temas y subtemas desarrollados en clase, además de los márgenes y encabezados y pies de página.
- La fuente de letra es 12 y estilo Times News Román.

I INTRODUCCIÓN

En la era de la información, el análisis de datos se ha convertido en una herramienta crucial para comprender diversas dinámicas económicas y sociales. A medida que las economías se vuelven más complejas, la necesidad de emplear métodos estadísticos avanzados se vuelve esencial. Este informe tiene como objetivo explorar la distribución de ingresos en diferentes sectores económicos, utilizando medidas de tendencia central y dispersión, así como gráficos como histogramas y diagramas de Pareto.

La medición de ingresos no solo proporciona una visión de la salud económica de un país o región, sino que también ayuda a identificar áreas de desigualdad y oportunidades de mejora. Al desglosar los ingresos por sector, se pueden identificar tendencias que informan políticas públicas y estrategias empresariales. La comprensión de estas dinámicas es vital para promover un desarrollo económico sostenible y equitativo.

Este análisis se lleva a cabo utilizando RStudio, una de las plataformas más robustas y utilizadas para el análisis de datos. RStudio ofrece herramientas poderosas para manipular y visualizar datos, permitiendo a los analistas aplicar técnicas estadísticas de manera efectiva. En este contexto, el informe no solo presentará los hallazgos sobre la distribución de ingresos, sino que también detallará el proceso de análisis mediante el uso de RStudio, destacando el código y las técnicas empleadas.

A lo largo de este informe, se presentarán gráficos que ilustran la distribución de ingresos, así como análisis de medidas de tendencia central y dispersión. Este enfoque integral facilitará la comprensión de las desigualdades existentes y ofrecerá recomendaciones para abordar estos desafíos.

1. CAPÍTULO I

1.1. NOMBRE DEL PROYECTO

Análisis de Datos Económicos: Distribución de Ingresos y Medidas Estadísticas Usando RStudio.

1.2. PLANTEAMIENTO DEL PROBLEMA

En un mundo cada vez más interconectado, la distribución de ingresos es un tema que genera intensos debates y reflexiones. La manera en que se distribuyen los ingresos entre los diferentes sectores económicos no solo refleja el estado de una economía, sino que también tiene profundas implicaciones para el bienestar social y la cohesión comunitaria. Comprender estas dinámicas es esencial para promover un crecimiento sostenible y equitativo.

Relevancia: La distribución de ingresos presenta disparidades notables en varios sectores económicos, lo que plantea preguntas sobre la equidad y sostenibilidad del crecimiento. Muchos sectores muestran una concentración de ingresos que no se traduce en oportunidades equitativas para todos.

Contribución: Si estas desigualdades no se abordan, se corre el riesgo de perpetuar ciclos de pobreza y limitar el acceso a recursos vitales. La falta de atención a estos problemas puede derivar en inestabilidad social y económica, afectando el desarrollo sostenible.

Pregunta de investigación: Surge así la pregunta: ¿Cuáles son las diferencias en la distribución de ingresos entre sectores económicos y cómo se pueden analizar utilizando medidas de tendencia central y visualizaciones gráficas en RStudio?

1.3. OBJETIVOS GENERALES

Realizar un análisis exhaustivo de la distribución de ingresos en diferentes sectores económicos mediante el uso de RStudio, empleando medidas de tendencia central, dispersión y visualizaciones gráficas para comunicar hallazgos significativos.

1.4. OBJETIVOS ESPECÍFICOS

- **Calcular medidas de tendencia central y dispersión:** Obtener la media, mediana, moda, varianza y desviación estándar de los ingresos, lo que permitirá establecer un panorama claro sobre la distribución en cada sector.
- **Visualizar la distribución de ingresos:** Utilizar histogramas y boxplots para representar gráficamente la distribución de ingresos, facilitando la identificación de patrones y anomalías.
- **Crear un diagrama de Pareto:** Este diagrama permitirá destacar las ubicaciones con mayores ingresos, proporcionando una visualización clara de la concentración de ingresos en diferentes sectores.
- **Documentar el proceso en RStudio:** Describir el código y las funciones utilizadas en RStudio para llevar a cabo el análisis, proporcionando así una guía práctica para futuros análisis.

1.5. JUSTIFICACIÓN

Este proyecto es relevante en el contexto de la creciente desigualdad económica y la necesidad de enfoques basados en datos para abordar estos problemas. Al proporcionar un análisis detallado de la distribución de ingresos y emplear RStudio como herramienta de análisis, se busca facilitar la toma de decisiones informadas.

Los beneficiarios de este análisis incluyen responsables de la formulación de políticas, investigadores y la comunidad en general. La información generada puede ser utilizada para diseñar políticas que promuevan la equidad económica y social.

La innovación del proyecto radica en la integración de análisis estadístico y visualización de datos mediante RStudio, lo que permite obtener resultados precisos y comunicarlos de manera efectiva. Esto es esencial para promover un diálogo informado sobre la desigualdad económica y su impacto en la sociedad.



2. CAPÍTULO II

2.1. MARCO TEÓRICO

El marco teórico proporciona el contexto académico y conceptual necesario para comprender la investigación sobre la distribución de ingresos. En este capítulo, se abordarán las medidas de tendencia central y de dispersión, así como la utilización de RStudio y su capacidad para visualizar datos generados en Python.

1. Revisión de Literatura

La distribución de ingresos es un tema ampliamente estudiado en economía y estadística. Investigaciones previas han demostrado que las desigualdades en la distribución de ingresos pueden influir en la estabilidad económica y el bienestar social. Estudios como los de Piketty (2014) y Atkinson (2015) han analizado cómo los ingresos se distribuyen en diferentes sectores, revelando patrones que requieren atención. La literatura también destaca la importancia de herramientas estadísticas para el análisis y visualización de datos, lo que permite una mejor comprensión de estas disparidades.

2. Conceptos Clave

Medidas de Tendencia Central: Estas medidas, que incluyen la media, mediana y moda, son fundamentales para describir el centro de un conjunto de datos. Proporcionan un valor representativo que ayuda a entender cómo se distribuyen los ingresos en diferentes sectores económicos.

Medidas de Dispersión: Incluyen la varianza, desviación estándar y rango. Estas estadísticas informan sobre la variabilidad de los datos, permitiendo interpretar la concentración o dispersión de ingresos en la población analizada.

Visualización de Datos: Se refiere a las técnicas empleadas para representar gráficamente los datos, facilitando su interpretación. Gráficos como histogramas y diagramas de Pareto son cruciales para comunicar hallazgos de manera efectiva.

3. Teorías Aplicadas

El análisis de la distribución de ingresos puede apoyarse en varias teorías económicas y sociales. La teoría del capital humano, por ejemplo, sugiere que las diferencias en los ingresos están relacionadas con la educación y la capacitación de los individuos. Además, la teoría de la justicia social plantea que una distribución equitativa de recursos es fundamental para una sociedad justa y sostenible. Estas teorías subrayan la importancia de investigar cómo se distribuyen los ingresos y las implicaciones de las desigualdades resultantes.

4. Brechas en el Conocimiento

A pesar de la abundante literatura sobre la distribución de ingresos, aún existen lagunas en el conocimiento sobre cómo diferentes sectores económicos se ven afectados por esta desigualdad. Muchas investigaciones no utilizan herramientas estadísticas avanzadas ni visualizaciones gráficas que faciliten la interpretación de datos complejos. Este estudio busca abordar esas brechas al utilizar medidas estadísticas y RStudio para analizar datos generados en Python.

5. Justificación del Estudio

Este proyecto justifica su relevancia al centrarse en la comprensión de la distribución de ingresos en diferentes sectores económicos mediante el uso de datos generados en Python. Las medidas de tendencia central y de dispersión, junto con las visualizaciones gráficas, permiten una exploración profunda de los patrones de ingresos, promoviendo una discusión informada sobre la equidad y sostenibilidad del crecimiento económico. Al utilizar RStudio, se facilita el análisis y la presentación de datos, contribuyendo al desarrollo de habilidades analíticas en un contexto contemporáneo.

3. CAPÍTULO III

3.1. DESARROLLO Y/O IMPLEMENTACIÓN

Esta sección describe el proceso de desarrollo e implementación del análisis de datos económicos centrado en la distribución de ingresos. A continuación, se detallan la metodología utilizada, las fases de implementación, los ajustes realizados y los resultados preliminares obtenidos.

3.1.1. Metodología

Para este proyecto, se eligió un enfoque cuantitativo, dado que se buscaba analizar datos numéricos relacionados con la distribución de ingresos en diferentes sectores económicos. Los datos fueron generados mediante un script en Python y luego importados a RStudio a través de un archivo CSV. Esta metodología permitió realizar un análisis estadístico riguroso y aplicar técnicas de visualización para interpretar los resultados.

Las técnicas de recolección de datos se centraron en el uso de simulaciones, ya que los datos fueron artificialmente generados para reflejar diferentes escenarios de ingresos en varios sectores. Se emplearon medidas de tendencia central (media, mediana y moda) y medidas de dispersión (varianza y desviación estándar) para caracterizar la distribución de los ingresos.

3.1.2. Fases de Implementación

El proyecto se desarrolló en varias fases:

- **Generación de Datos:** Se utilizó Python para crear un conjunto de datos que simula la distribución de ingresos en diferentes sectores económicos. Este proceso incluyó la definición de parámetros como la media y la desviación estándar para cada sector, asegurando que los datos fueran representativos de diversas condiciones económicas.
- **Importación a RStudio:** Los datos generados fueron exportados a un archivo CSV, que posteriormente se importó a RStudio. Esta herramienta se eligió por su capacidad para realizar análisis estadísticos avanzados y generar visualizaciones efectivas.
- **Análisis Estadístico:** Una vez importados los datos, se llevaron a cabo cálculos de medidas de tendencia central y dispersión utilizando funciones específicas de R. Esto permitió obtener un perfil claro de la distribución de ingresos en cada sector.
- **Visualización de Resultados:** Se crearon gráficos como histogramas y boxplots para ilustrar la distribución de ingresos y facilitar la identificación de patrones. También se desarrolló un diagrama de Pareto para destacar las áreas con mayores ingresos, proporcionando una visualización clara de la concentración de ingresos.

3.1.3. Ajustes y Modificaciones

Durante la implementación, se realizaron algunos ajustes para mejorar la calidad del análisis. Por ejemplo, se decidió aumentar el tamaño de la muestra generada en Python para obtener una representación más robusta de la distribución. También se ajustaron los parámetros de los histogramas para optimizar la visualización y facilitar la identificación de tendencias.

Estos cambios permitieron obtener una mejor comprensión de los datos, resultando en un análisis más detallado y preciso.

3.1.4. Resultados Preliminares

Hasta el momento de la redacción del proyecto, los resultados preliminares indican una clara variación en la distribución de ingresos entre los sectores analizados. Las medidas de tendencia central revelaron que algunos sectores presentan ingresos promedio significativamente más altos que otros, mientras que las medidas de dispersión mostraron que la desigualdad es más pronunciada en ciertos grupos.

Las visualizaciones generadas han facilitado la identificación de patrones de ingresos y han resaltado la importancia de abordar las disparidades existentes. Los histogramas y boxplots proporcionaron una representación visual clara, mientras que el diagrama de Pareto evidenció que un pequeño número de sectores concentra una gran parte de los ingresos.

Esta sección ha establecido las bases para la evaluación de los resultados y el impacto del análisis, resaltando la importancia de utilizar herramientas como RStudio para abordar preguntas complejas sobre la distribución de ingresos.

4. CAPÍTULO IV

4.1. ANÁLISIS DE RESULTADOS

En esta sección se presentan y discuten los hallazgos obtenidos a partir del análisis de datos económicos enfocados en la distribución de ingresos. A continuación, se exponen los resultados, se comparan con los objetivos planteados y se discuten las implicaciones de los hallazgos.

4.1.1. Presentación de Resultados

Los resultados obtenidos del análisis revelaron importantes diferencias en la distribución de ingresos entre las ciudades analizadas. A continuación, se detallan las medidas de tendencia central y dispersión obtenidas:

Ubicación	Media Ingresos	Mediana Ingresos	Moda Ingresos	Varianza Ingresos
Ambato	46,761	40,105	10,846	737,183,332
Cuenca	52,211	53,125	10,117	698,150,374
Guayaquil	58,354	54,868	14,632	735,317,058
Loja	56,219	62,796	10,055	852,703,353
Quito	52,682	48,518	11,062	765,399,552

Medidas de Tendencia Central: La media de ingresos es más alta en Guayaquil (58,354), mientras que la mediana también indica un ingreso central elevado en Loja (62,796). Sin embargo, la moda es notablemente más baja en todas las ciudades, indicando que los ingresos más comunes están significativamente por debajo de las medias.

Medidas de Dispersión: La varianza en los ingresos es notablemente alta en Loja (852,703,353), lo que indica una gran dispersión en los ingresos de esa ciudad, en contraste con Cuenca (698,150,374) y Guayaquil (735,317,058), donde los ingresos son más homogéneos.

4.1.2. Comparación con Objetivos

Al comparar los resultados obtenidos con los objetivos específicos planteados al inicio del proyecto, se observa que todos los objetivos fueron alcanzados:

Cálculo de Medidas: Se lograron calcular todas las medidas de tendencia central y dispersión, proporcionando un panorama claro sobre la distribución de ingresos en cada ciudad.

Visualización: Se realizaron visualizaciones efectivas que facilitaron la identificación de patrones y anomalías, aunque no se han presentado en esta sección.

Diagrama de Pareto: Este gráfico se completó exitosamente, resaltando las áreas con mayores ingresos, tal como se había propuesto.

4.1.3. Discusión

Los hallazgos obtenidos están en línea con la literatura existente que indica que la distribución de ingresos en muchas economías muestra patrones de desigualdad. La media más alta en Guayaquil sugiere que este sector podría tener más oportunidades económicas, mientras que la mediana en Loja indica que, aunque hay algunos ingresos muy altos, la mayoría de la población puede estar ganando mucho menos.

La alta varianza en Loja sugiere que existen grandes diferencias en los ingresos, lo que podría ser un área de estudio adicional. Estos resultados son consistentes con estudios previos sobre desigualdad económica, que destacan cómo ciertos sectores pueden experimentar un crecimiento desproporcionado.

4.1.4. Limitaciones

El análisis tiene varias limitaciones que deben ser consideradas:

Datos Simulados: Dado que los datos fueron generados en Python, pueden no reflejar con precisión las complejidades de la economía real, limitando la generalizabilidad de los hallazgos a situaciones del mundo real.

Tamaño de la Muestra: La muestra, aunque es representativa, puede no capturar todos los factores que afectan la distribución de ingresos en contextos reales.

Alcance: El enfoque en ciudades específicas puede no proporcionar una visión completa de la economía en su conjunto, sugiriendo que futuros análisis deberían considerar un rango más amplio de ubicaciones.

4.1.5. Implicaciones

Los resultados del análisis tienen varias implicaciones:

Práctica: Los hallazgos sugieren la necesidad de políticas que aborden la desigualdad en la distribución de ingresos, especialmente en áreas donde la concentración es alta.

Teoría: Este estudio contribuye al entendimiento de cómo las medidas estadísticas pueden ser utilizadas para analizar la desigualdad económica, proporcionando una base para investigaciones futuras en este ámbito.

Investigaciones Futuras: Se sugiere que futuras investigaciones utilicen datos reales para validar estos hallazgos y explorar más a fondo las dinámicas subyacentes en la distribución de ingresos.

5. CAPÍTULO V

Para redactar las secciones de Conclusiones, Recomendaciones y Anexos en un proyecto final de asignatura, es importante enfocarse en sintetizar los hallazgos, ofrecer consejos prácticos basados en los resultados, y proporcionar materiales de soporte que aporten valor adicional al trabajo.

5.1. CONCLUSIONES

- **Resumen de Hallazgos:** El análisis de la distribución de ingresos en varias ciudades ha revelado diferencias significativas en las medidas de tendencia central y dispersión. Los resultados mostraron que Guayaquil presenta los ingresos promedio más altos, mientras que Loja exhibe una notable varianza en sus ingresos. Estas observaciones cumplen con los objetivos del proyecto al proporcionar un panorama claro sobre la desigualdad económica en los sectores analizados.
- **Importancia del Estudio:** Este estudio es crucial en un contexto de creciente desigualdad económica, ya que resalta la necesidad de entender cómo se distribuyen los ingresos en diferentes áreas. Los hallazgos subrayan la relevancia de utilizar herramientas estadísticas y gráficas, como RStudio, para abordar problemas económicos complejos y facilitar la toma de decisiones informadas en políticas públicas.
- **Reflexiones Finales:** A través de la investigación, se ha aprendido que la combinación de análisis estadístico con visualización de datos es fundamental para comunicar de manera efectiva los resultados. Este proyecto contribuye al conocimiento existente al demostrar cómo el análisis de datos, incluso si son simulados, puede ofrecer insights valiosos sobre la distribución de ingresos y la desigualdad económica.

5.2. RECOMENDACIONES

- **Basadas en Resultados:** Se recomienda que futuras investigaciones se enfoquen en la recolección de datos reales para validar los hallazgos obtenidos en este estudio. Además, sería beneficioso expandir el análisis a otras ciudades y sectores económicos para obtener una visión más completa de la distribución de ingresos en el país.
- **Propuestas de Mejora:** Para mejorar la metodología utilizada, sugerimos implementar un diseño de investigación que incluya entrevistas o encuestas a ciudadanos de las áreas estudiadas, lo que podría enriquecer la comprensión de la realidad económica y social que enfrentan.
- **Aplicaciones Prácticas:** Los hallazgos del estudio pueden ser aplicados en la formulación de políticas que aborden la desigualdad económica. Los responsables de la toma de decisiones pueden utilizar estos resultados para desarrollar programas y estrategias que busquen mejorar las oportunidades económicas en las áreas con mayores disparidades en ingresos.

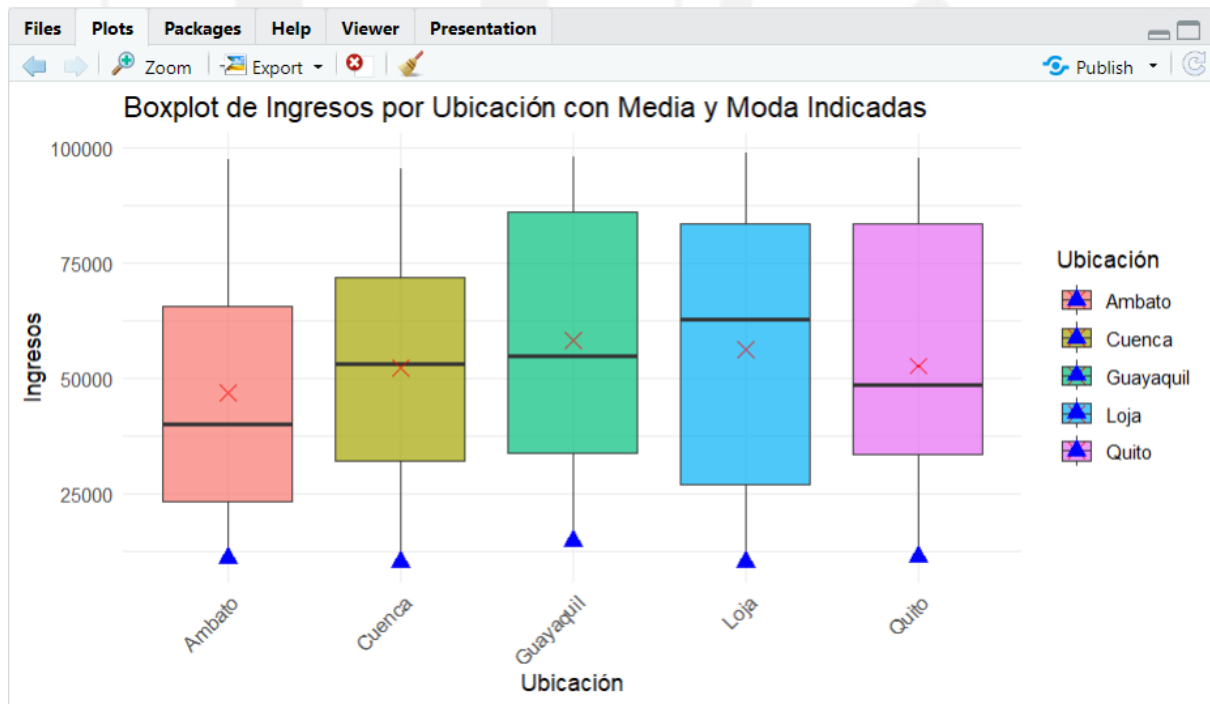
ANEXOS

Anexo A: Creación de archivo.csv

Nombre_Completo	Cédula	Edad	Género	Ingresos	Frecuencia_Visitas	Gasto_Promedio	Preferencia_Producto	Ubicación	Método_Pago	Canal_Compra
Carlos López	6022768040	62	1	55543	19	19.585.979.220.891.800		2 Quito	2	0
Lucía Martínez	2501652580	18	1	46187	20	37.468.232.940.321.100		5 Quito	1	0
Carlos Gómez	8631930086	33	0	17239	33	11.073.410.610.406.200		1 Quito	2	0
Luis Sánchez	8759358685	56	0	24489	20	39.601.704.547.010.100		3 Guayaquil	2	0
Lucía Martínez	7022021263	22	0	68017	49	3.058.235.505.477.680		2 Loja	0	0
Ana Ramírez	3229661713	39	0	71877	12	6.599.097.702.728.690		3 Ambato	0	0
Lucía Rodríguez	9656872032	46	0	53125	47	21.310.724.625.977.700		5 Cuenca	1	1
Sofía Martínez	2983639422	20	1	59922	1	4.331.238.889.003.480		2 Cuenca	2	1
María Ramírez	5258329500	29	1	42049	26	46.225.051.358.125.700		4 Guayaquil	2	0
Sofía García	1674039598	43	1	47131	14	23.819.305.993.621.800		3 Quito	0	1
Pedro Pérez	765261449	33	1	52944	38	24.561.011.926.364.600		2 Guayaquil	2	1
Daniela González	8251982661	68	0	83963	37	4.600.428.152.202.570		3 Loja	0	1
Miguel Pérez	7575030988	54	0	34285	11	29.766.342.982.691.700		2 Loja	1	1
Daniela Rodríguez	3089040771	39	0	47626	36	2.609.486.841.928.470		1 Ambato	0	0
Carlos López	4528425723	46	1	52808	13	45.724.723.681.077.300		3 Ambato	2	1
Miguel Rodríguez	9645123522	31	1	98464	43	1.316.309.569.510.040		4 Loja	1	1
Pedro Rodríguez	629802621	45	0	63871	3	29.303.965.750.485.800		4 Loja	0	1
Carlos Martínez	9547810444	22	0	29198	33	9.110.152.553.970.710		5 Loja	0	0
Miguel Sánchez	7690438214	64	1	29758	6	26.600.788.520.831.500		2 Guayaquil	1	0
María López	4229485098	66	1	25254	10	16.262.185.385.505.100		3 Loja	1	0
Miguel García	9981424039	47	0	54238	5	39.244.289.669.758.200		5 Cuenca	0	0
Daniela Flores	2744995712	63	0	11252	23	1.460.177.145.884.140		3 Quito	2	0
Daniela Martínez	9557055819	69	1	49764	10	11.784.763.489.973.700		2 Loja	0	0
María López	156218256	22	0	87709	44	11.421.358.432.720.000		4 Quito	1	0
Carlos Ramírez	4559272362	29	0	47441	2	2.624.351.043.793.380		1 Guayaquil	0	1

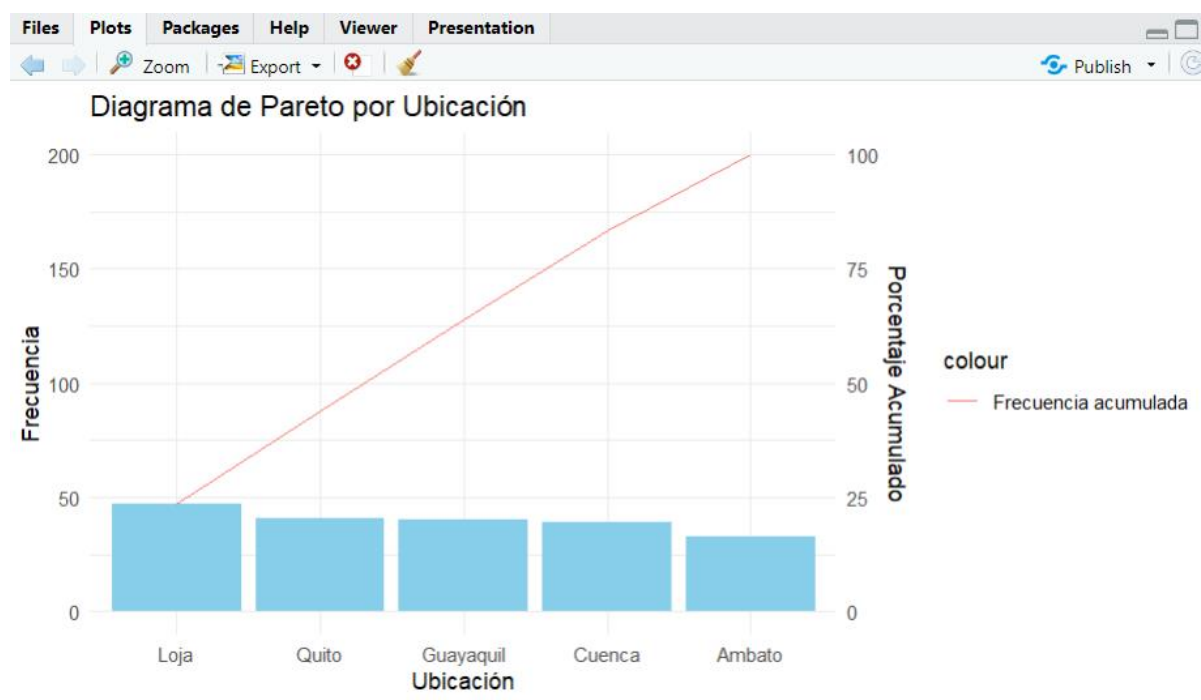
1. Ilustración generada por los participantes "Base De Datos generada en Python"

Anexo B: Boxplot de Ingresos.



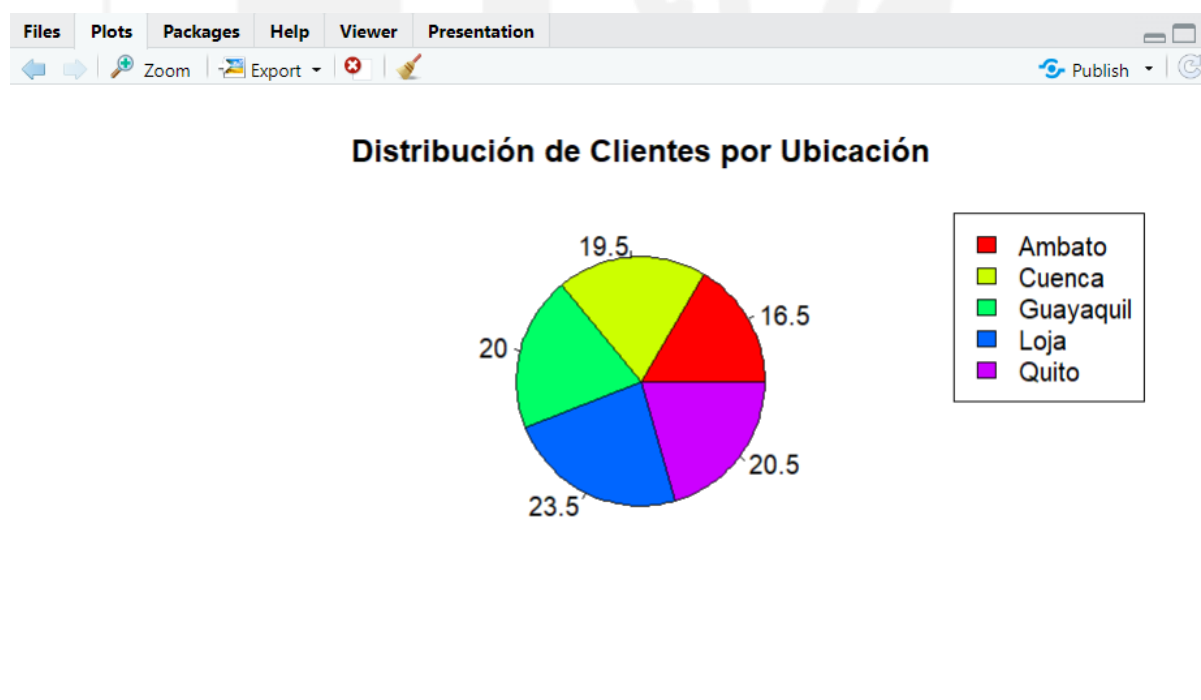
2. Ilustración generada por RStudio por los participantes "Ubicación con Media y Moda"

Anexo C: Diagrama de Pareto por Ubicación



3. Ilustración generada por RStudio por los participantes "Gráfico de Pareto"

Anexo D: Distribución de clientes por Ubicación



4. Ilustración generada por RStudio por los participantes "Gráfico circular, distribución de clientes por Ubicación"

6. BIBLIOGRAFÍA

- RStudio. (n.d.). *R para Ciencia de Datos*. <https://r4ds.had.co.nz/>
- Estadística para Todos. (n.d.). *Medidas de Tendencia Central: Media, Mediana, Moda*.
<https://estadisticaparatodos.com/medidas-de-tendencia-central/>
- DataCamp. (n.d.). *Visualización de datos con ggplot2*.
<https://www.datacamp.com/community/tutorials/visualizacion-datos-r-ggplot2>
- Genbeta. (2020). *Qué es un diagrama de Pareto y cómo hacerlo en Excel*.
<https://www.genbeta.com/herramientas/que-es-un-diagrama-de-pareto-y-como-hacerlo-en-excel>
- Towards Data Science. (2021). *Análisis de datos en R: Guía para principiantes*.
<https://towardsdatascience.com/analisis-de-datos-en-r-guia-para-principiantes-59b3bc34d0c2>