



Minería de Datos

Facultad de Matemáticas, Ciencia de Datos

Mag. Albert Ferney Montenegro Vargas

Análisis Exploratorio de Datos de Ventas en Nueva York.

Salomé Castro Gómez

Samuel Fernando Monroy Herrera

Danna Valbuena Plazas

*“La imaginación es más importante que el conocimiento. El conocimiento es limitado; la imaginación rodea el mundo” - **Albert Einstein***

14 de mayo de 2024

Bogotá D.C, Colombia

INDICE

1.	Resumen.....	3
2.	Introducción	3
3.	Contexto	4
4.	Diccionario de variables	5
5.	Tipología de variables.....	7
6.	Método de filtro I	7
7.	Análisis Univariado.....	8
7.1.	Barrio.....	8
7.2.	Clase de Construcción	9
7.3.	Clase de Impuesto Actual	10
7.4.	Borough.....	10
7.5.	Year Built	10
7.6.	Sale Price.....	11
8.	Métodos de filtro II y III	11
9.	Imputación.....	12
10.	Análisis Bivariado	12
10.1.	Unidades Residenciales contra Distrito.....	12
10.2.	Unidades Comerciales contra Distrito.....	13
10.3.	Unidades Comerciales contra Tipo de Construcción.....	13
10.4.	Unidades Residenciales contra Tipo de Construcción.....	14
10.5.	Año de Construcción contra Distrito	14
10.6.	Año de Construcción contra Tipo de Construcción.....	15
10.7.	Año de Construcción contra Precio de Venta.....	16
10.8.	Distrito contra Fecha de Venta	17
10.9.	Distrito contra Tipo de Construcción	18
10.10.	Distrito contra Clase Fiscal	18
10.11.	Distrito contra Precio de Venta	19
10.12.	Clase fiscal contra Precio de Venta.....	19
10.13.	Tipo de construcción contra Precio de Venta.....	20
11.	Resumen de las pruebas del Análisis Bivariado.....	22
12.	Conclusiones	23
13.	Bibliografía	24

1. Resumen

Nueva York es una ciudad de especial interés para los inversores del sector inmobiliario, proporcionando amplia variedad de ofertas según las necesidades e intereses del individuo. Mediante la exploración y análisis de la base de datos “NYC Property Sales” sobre la venta de propiedades inmobiliarias en Nueva York se detectaron múltiples factores que tienen efecto sobre el precio de venta; entre ellos, se detectó que el distrito cumple un rol fundamental en el valor de la edificación, ya que cada uno de los 5 *borough* posee un desarrollo e infraestructura urbanística diferente, haciendo que ciertas locaciones sean más apropiadas para la vivienda o el establecimiento comercial. Además, la estacionalidad y la fecha causa fluctuaciones en los precios, siendo su cumbre en meses asociados a la primavera y el invierno. Para realizar el análisis fue fundamental una profunda investigación sobre el contexto histórico del mercado inmobiliario, por tal motivo, se pudo evidenciar el auge de diversos tipos de edificios desde inicios del siglo XX hasta la actualidad, comparándolos con diversos fenómenos socioeconómicos y demográficos.

Finalmente, se detallan una serie de conclusiones que brindan un panorama general sobre los hallazgos realizados en este estudio, con el objetivo de dar un diagnóstico sobre el mercado inmobiliario neoyorquino y ser de utilidad para la resolución de fenómenos críticos como las inversiones, la gentrificación y la crisis de vivienda en Estados Unidos.

2. Introducción

La minería de datos es una herramienta fundamental para extraer información valiosa y dar conocimiento idóneo en gran volumen de datos, esto nos da paso para emplear técnicas analíticas y modelos aplicables a la necesidad correspondiente, de este mismo modo descubrir patrones, tendencias y relaciones ocultas.

En el contexto del mercado inmobiliario de la ciudad de Nueva York, el cual es un sector demandado y altamente competitivo, el énfasis sobre esta disciplina desempeña un papel fundamental. Con un año completo de registros de transacciones inmobiliarias, se presenta una oportunidad para aplicar las técnicas anteriormente mencionadas, por consiguiente, obtener información valiosa para comprender su comportamiento e igualmente darles respuesta a las siguientes preguntas:

- *¿Qué se puede descubrir sobre el mercado inmobiliario neoyorquino observando un año de registros de transacciones sin procesar?*
- *¿Se pueden detectar tendencias en el mercado o construir un modelo que prediga el valor de venta en el futuro?*

El estudio se llevará a cabo mediante EDA, en el cual se aborda análisis univariado y bivariado, visualizaciones, observación de correlaciones, y en última instancia la construcción de un modelo predictivo. Estos hallazgos pueden brindar información valiosa sobre la influencia de factores como la ubicación, el tipo de propiedad, el comportamiento de precio y observar una comprensión profunda de las tendencias y patrones que impulsan el mercado.

3. Contexto

“NYC Property Sales” es una base de datos publicada por la administración de Nueva York que contiene información sobre diversas transacciones inmobiliarias realizadas entre 2016 y 2017.

El mercado inmobiliario en Nueva York es uno de los más robustos y atractivos en el mundo dado la amplia oferta de propiedades en diversas locaciones y ambientes, permitiendo a miles de personas escoger entre apartamentos de lujo, condominios, casas, edificios familiares, entre otros...

En esta medida, conocer los movimientos y rentabilidad de este mercado es fundamental para analizar tendencias, prever cambios e invertir basándose en lo que dicen los datos, el contexto y la historia.

Ahora bien, en los últimos años Nueva York ha atravesado fenómenos de gran impacto tales como la gentrificación, donde multinacionales e individuos con alto poder adquisitivo valorizan zonas de bajos recursos; obligando a miles de habitantes a abandonar sus hogares al ser incapaces de satisfacer el coste de vida. Igualmente, se estudiarán otros fenómenos críticos del mercado financiero en Nueva York.

Teniendo en cuenta lo anteriormente mencionado se realizó un análisis profundo de la base de datos siguiendo una serie de pasos y metodologías que se presentan en las siguientes secciones.

4. Diccionario de variables¹

1. **Building Class Category** (*Categoría de clase de construcción*): Contiene el tipo de construcción a la cual pertenece cada propiedad. Hay un total de 47 tipos de construcciones.
2. **Building Class at Present** (*Clase de construcción en la actualidad*): Describe el uso constructivo de cada propiedad.
3. **Building Class at Time of Sale** (*Clase de construcción en el momento de la venta*): Describe el uso constructivo de una propiedad.
4. **Tax Class at Time of Sale** (*Clase de impuesto en el momento de la venta*): Contiene la clase de impuesto de cada propiedad.
 - **Clase 1:** Propiedades residenciales de hasta tres unidades, terrenos baldíos para uso residencial y la mayoría de los condominios que no tengan más de tres pisos.
 - **Clase 2:** Incluye todas las demás propiedades que son principalmente residenciales, como cooperativas y condominios.
 - **Clase 3:** Incluye inmuebles con equipos de propiedades de compañías de gas, teléfono o electricidad.
 - **Clase 4:** Incluye todos los demás inmuebles no incluidos anteriormente en las clases 1, 2 y 3, como oficinas, fábricas, almacenes, garajes entre otros.
5. **Lote:** Subdivisión de un bloque fiscal y representa la ubicación única de la propiedad.
6. **Borough** (*Distrito*): Corresponde al distrito de Nueva York en el que se ubica la propiedad. Puede tomar cinco valores:
 - Manhattan
 - Queens
 - Staten Island
 - Brooklyn
 - Bronx
7. **Neighborhood** (*Barrio*): Corresponde al barrio particular en el que se ubica la propiedad. En total son **254** Barrios que se presentan en toda la base de datos
8. **Zip Code** (*Código Postal*): Código ZIP estándar correspondiente a la zona donde se ubica la propiedad
9. **Land Square Feet** (*Pies Cuadrados Netos*): Área útil en pies cuadrados de la propiedad. Corresponde únicamente al área habitable.
10. **Gross Square Feet** (*Pies Cuadrados Brutos*): Área útil y no útil en pies cuadrados de la propiedad. Corresponde al área habitable y no habitable (Ocupado por escaleras, muros entre otros)
11. **Year Built** (*Año de Construcción*): Año en el que fue construida la propiedad
12. **Block** (*Bloque fiscal*): Identificador único para ciertas zonas de la ciudad que facilita la recaudación fiscal
13. **Address** (*Dirección*): Identificador único para cada propiedad según la zona y las calles en que se ubica.
14. **Easement** (*Servidumbre*): Es un derecho, como un derecho de paso, que permite a una entidad hacer uso limitado de la propiedad inmueble de otra. Ejemplo: Las vías del ferrocarril que atraviesan una parte de otra propiedad.
15. **Residential Units** (*Unidades Residenciales*): Número total de unidades en la propiedad, puede incluir apartamentos, condominios, viviendas de una, dos y tres familias, entre otras.
16. **Commercial Units** (*Unidades Comerciales*): Número total de unidades destinadas para el uso comercial en las propiedades, contiene locales, hoteles, oficinas entre otros.

¹ (NYC, s.f.)

17. **Total Units** (*Unidades totales*): Es la suma de unidades residenciales y unidades comerciales.
18. **Sale Date** (*Fecha de Venta*): Es la fecha en la cual se realizó la venta de la propiedad.
19. **Apartment Number** (*Número de apartamento*): Número característico de cada apartamento.
20. **Tax Class at Present** (*Clase de impuesto actual*): En el glosario no hay información acerca de esta variable y contiene valores como: '1', '2', '3', '4', '2A', '2C', '1A', '1B', '2B', '1C' se infiere que hace referencia a lo mismo que la clase fiscal en el momento de la venta
21. **Sale Price** (*Precio de Venta*): Coste pagado por la propiedad. **Variable objetivo.**

5. Tipología de variables

Tabla 1. Tipología de Variables

<i>Variables Categóricas</i>	<i>Variables Numéricas</i>	<i>Meta Variables</i>	<i>Fechas</i>
Building Class category	Land Square Feet	Lote	Sale Date
Building Class at Present	Gross Square Feet	Block	Year Built
Building Class at Time of Sale	Year Built	Zip Code	-
Tax Class at Time of Sale	Residencial Units	Address	-
Borough	Commercial Units	Apartment Number	-
Neighborhood	Total Units	-	-
Tax Class at Present	Sale Date	-	-
-	Sale Price	-	-

Easement

La variable no posee ningún dato; es decir, no hay forma de describirla ni estudiarla.

6. Método de filtro I

El primer método de filtro utilizado consistió en descartar las variables que contenían identificadores únicos y se encuentran en la Tabla 1 en la columna de meta variables. Cabe recalcar que se descartó la variable Easement debido a que no posee ningún valor. En esta medida, las *features* utilizadas en la siguiente sección fueron:

- Building Class category
- Building Class at Present
- Building Class at Time of Sale
- Tax Class at Time of Sale
- Borough
- Neighborhood
- Tax Class at Present
- Land Square Feet
- Gross Square Feet
- Year Built
- Residencial Units
- Commercial Units
- Total Units
- Sale Date
- Sale Price

7. Análisis Univariado

Se iniciará dando un vistazo general al comportamiento individual de cada una de las variables, determinando la entropía y los valores faltantes (para el caso de las variables categóricas) o la normalidad, varianza y porcentaje de faltantes (para el caso de las variables numéricas). Los resultados se presentan en las Tablas 1 y 2 respectivamente.

Tabla 2. Análisis Univariado de Variables Categóricas

Variables Categóricas	Entropía	Porcentaje de Faltantes
Building Class Category	3.464	0%
Building Class at Present	4.617	0%
Building Class at Time of Sale	4.580	0%
Borough	2.149	0%
Neighborhood	7.367	0%
Tax Class at Present	1.930	0%
Tax Class at Time of Sale	1.305	0%

Tabla 3. Análisis Univariado de Variables Numéricas

Variables Numéricas	Normalidad ²	Coefficiente de Variación	Valores Faltantes
Sale Price	No	893.50 %	17.22%
Residential Units	No	825.62 %	0%
Commercial Units	No	4501.57 %	0%
Total Units	No	843.53 %	0%
Land Square Feet	No	1065 %	43%
Gross Square Feet	No	865 %	46%

Para determinar el comportamiento de variables en específico se analizaron a profundidad cada una de ellas con el fin de observar particularidades de los datos. A continuación, se presentan los hallazgos más relevantes.

7.1. Barrio

Todas las propiedades tienen algún barrio asignado en la base de datos. Con el fin de obtener más información al respecto, se realizó una investigación sobre los 5 barrios con más ventas registradas:

² Se utilizaron las pruebas de Kolmogorov, Shapiro y D'Agostino obteniendo en todos los casos un P valor de 0. Se rechaza H_0 (los datos son normales) con una significancia de 0.05

- **Flushing North:** Barrio donde se ubica el cuarto distrito financiero más importante de Nueva York. Es uno de los más multiculturales del estado, posee alta influencia asiática y en los últimos años ha sido altamente afectado por la gentrificación.³
- **Upper East Side/ Upper West Side:** Barrios reconocidos por ser las zonas más exclusivas y lujosas del estado, siendo lugar de residencia de los individuos con mayor poder adquisitivo.⁴
- **Bed-Stuy:** Barrio conocido por ser el hogar de la comunidad afroamericana de Brooklyn y muchas otras personas desplazadas de Harlem, Nueva York. En los últimos años ha sufrido un abrupto proceso de gentrificación.⁵
- **Midtown East:** Barrio de Manhattan que alberga importantes centros financieros y económicos como Times Square Building y La Gran Manzana. En los últimos años ha aumentado el número de personas sin hogar debido a los altos precios del mercado inmobiliario en la zona.⁶

7.2. Clase de Construcción

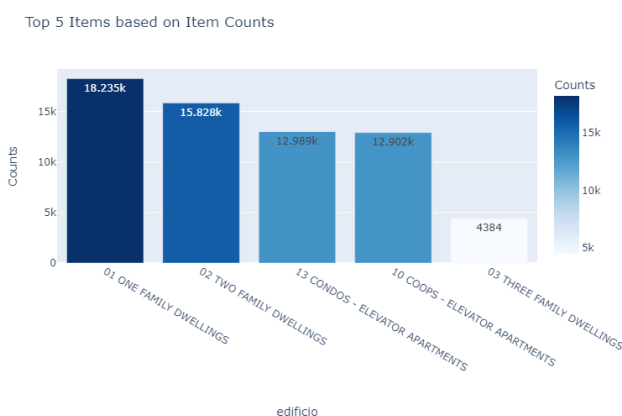


Ilustración 1. 5 edificios más comunes

En Ilustración 1 se observan los 5 tipos de edificios que más se construyen, entre ellos resalta el tipo *One family Dwellings*. Se detallarán cada uno de ellos a continuación:

- **One Family Dwellings:** Viviendas unifamiliares, en las que habitan solo una familia y no se encuentran más de cinco inquilinos o huéspedes.
- **Two Family Dwellings:** Viviendas bifamiliares en las cuales viven dos familias independientes entre sí.
- **Condos – Elevator Apartments:** Condominios los cuales cuentan con elevador privado.
- **Coops- Elevator Apartments:** Apartamentos en los que los inquilinos comparten el ascensor y son cooperativos.

³ (Guiajando)

⁴ (Nuevayork.es)

⁵ (Guiajando)

⁶ (Cifuentes, 2022)

- **Three family Dwellings:** Edificaciones en las que viven tres familias independientes entre sí, cada residencia con su propia entrada. Ejemplo: Edificios triplex.

7.3. Clase de Impuesto Actual

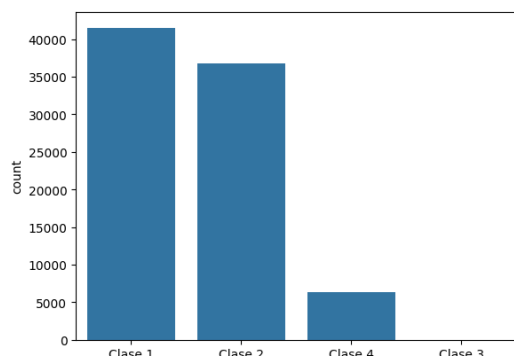


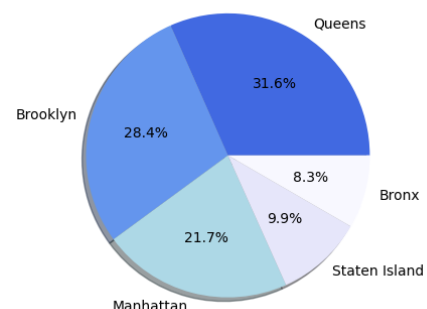
Ilustración 2. Frecuencias de las clases fiscales

En Ilustración 2 se evidencia que los impuestos se distribuyen mayormente entre la clase uno y dos. Se infiere que en Nueva York la mayoría de las edificaciones son de tipo residencial y muy pocas son oficinas, fábricas, almacenes, garajes entre otros. Además, son casi nulas las edificaciones de tipo 3 tales como compañías de teléfono, gas o electricidad.

7.4. Borough

Todas las propiedades registradas tienen uno de los 5 distritos asignados; además, en Ilustración 3 se puede notar que la gran parte de los datos se ubican en Brooklyn o Queens (corresponden al 60% del total). Por su parte, Bronx y Staten Island son los que menos actividad comercial poseen, reuniendo juntos apenas el 18.2% del total.

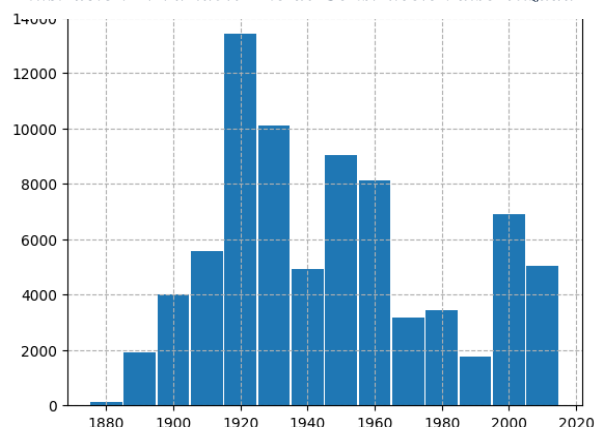
Ilustración 3. Porcentaje de registros por distrito



7.5. Year Built

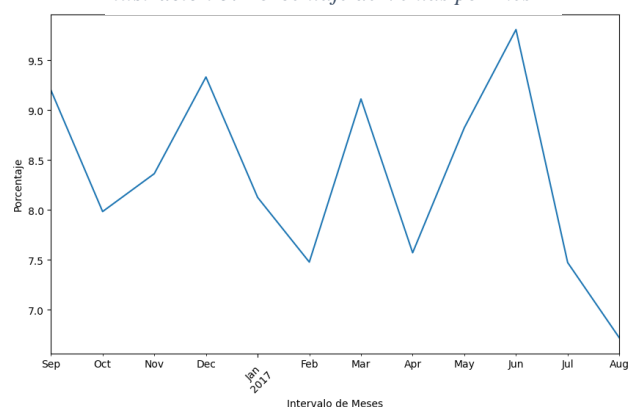
Inicialmente se detectan registros marcados en 0 y uno en el año 1111 (cuando aún no existía EE. UU.), se considera que ambos casos corresponden a errores y se marcan como faltantes (8%). La distribución de la variable revela picos anómalos en los años múltiplos de 5; dado que se desconoce la naturaleza de este aspecto, se decide discretizar la variable en décadas (y todos los años inferiores a 1880 se almacenan en una sola categoría) tal como se muestra en Ilustración 4

Ilustración 4. Variable Año de Construcción discretizada



Hay picos en los veinte, correspondientes al proceso de industrialización y globalización que llevó a EE. UU pasó a ser una potencia mundial, posteriormente hubo una caída en los treinta y cuarenta causada por La Gran Depresión. En los últimos años hay una tendencia al alza.

Ilustración 5. Porcentaje de Ventas por mes



7.6. Sale Price

En la Ilustración 5 se puede observar el comportamiento de las fechas respecto al número de ventas inmobiliarias: hay picos entre los meses de noviembre, diciembre, marzo y mayormente en junio. También se evidencian valles en los meses de octubre, febrero, abril y una decadencia en el mes de agosto. Esto nos da una visión global del comportamiento por fechas más frecuentes.

8. Métodos de filtro II y III

Para esta parte utilizamos diferentes criterios para filtrar las variables entre ellos se decidió no incluir *Total Units* debido a que esta información representa la suma de unidades comerciales y residenciales, lo cual resultaba redundante. Del mismo modo, se optó por descartar las variables *Building Class at Present* y *Building Class at Time of Sale*, ya que estas prácticamente duplicaban la información proporcionada por *Building Class Category*. Como tercer método de filtro, se excluyó *Tax Class at Present* debido a su similitud con *Tax Class at Time of Sale*. Por último, se decidió eliminar *Land Square Feet* y *Gross Square Feet* debido a que más del 40% de los valores estaban ausentes, lo que comprometía significativamente la integridad de los datos.

9. Imputación

Teniendo en cuenta las 9 variables filtradas y el análisis univariado respectivo a cada una de ellas, se identifica que *Año de Construcción*, *Tipo de Construcción Actual* y *Precio de Venta* poseen valores faltantes; por tanto, se decide implementar un algoritmo de aprendizaje automático para realizar imputación y facilitar el análisis bivariado.

Se hizo uso de K – NN; sin embargo, dado el tamaño de la base de datos y la carencia de gran capacidad de cómputo, se decidió probar únicamente con $k = \sqrt{(Total\ de\ datos)}$, el cual, según diversos autores puede ser una buena aproximación al comportamiento ideal.⁷

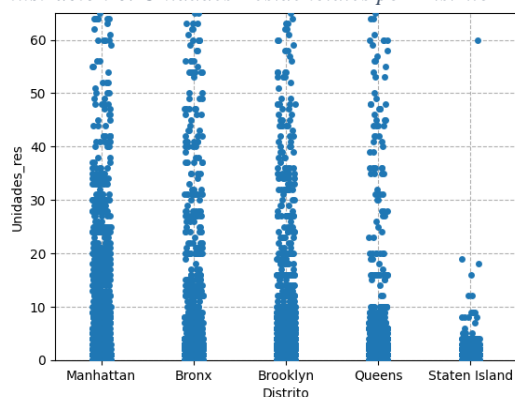
Tras realizar la implementación del modelo, la base de datos con las variables filtradas quedó sin ningún valor faltante.

10. Análisis Bivariado

10.1. Unidades Residenciales contra Distrito

Se aplicó la prueba Kruskal – Wallis, obteniendo un P – valor de 0 y se evidencia que existe dependencia entre ellas.

Ilustración 6. Unidades Residenciales por Distrito



Particularmente se observa que Staten Island tiende a tener menos cantidad de unidades residenciales en sus propiedades, contrario al caso de Manhattan, el cual, de entre todos los distritos, presenta mayor acumulación en valores altos, seguido de Brooklyn, Bronx y Queens. Adicionalmente se identificaron atípicos, determinando una gran cantidad en todos los distritos.

Tabla 4. *Atípicos de Unidades Residenciales por Distrito*

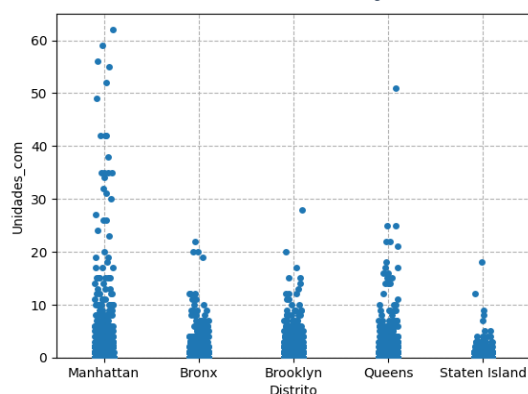
	Número de Atípicos
Bronx	6816.0
Manhattan	18189.0
Brooklyn	6568.0
Queens	19212.0
Staten Island	122.0

Dado que corresponden a atípicos naturales (causado porque efectivamente hay edificios con muchas residencias como hoteles, apartamentos de muchos pisos, entre otros) Se decide no aplicar ningún procedimiento sobre los datos.

⁷ (Universidad Nacional de Cuyo)

10.2. Unidades Comerciales contra Distrito

Ilustración 7. Unidades Comerciales por Distrito



Análogo a la relación anterior, se aplicó Kruskal – Wallis obteniendo un P – valor de 0; es decir, existe dependencia entre las variables.

Mediante un stripplot se observa el comportamiento por distrito, detectando, una vez más, una abrumadora diferencia entre Manhattan y Staten Island, siendo el primero, el distrito con mayor movimiento y actividad comercial.

Tabla 5. Atípicos de Unidades Comerciales por Distrito

	Número de Atípicos
Bronx	7049.0
Manhattan	18306.0
Brooklyn	24047.0
Queens	26736.0
Staten Island	8410.0

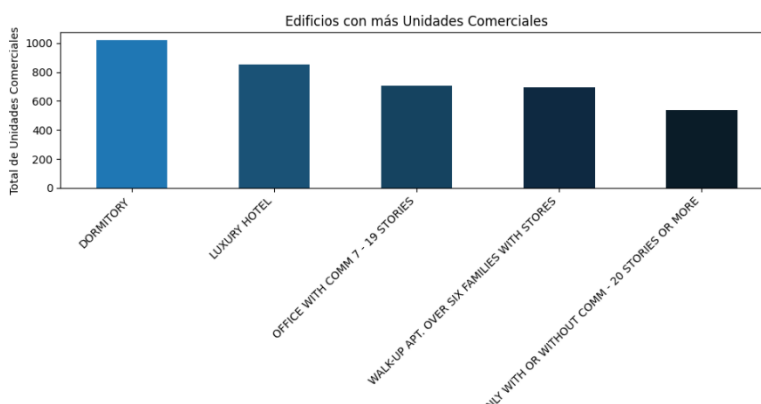
Asimismo, se identificaron atípicos obteniendo los siguientes resultados.

Igual que en el caso anterior, estos atípicos se consideran naturales dado que obedecen al contexto propio de la base de datos, por lo que no se aplica ningún procedimiento.

10.3. Unidades Comerciales contra Tipo de Construcción.

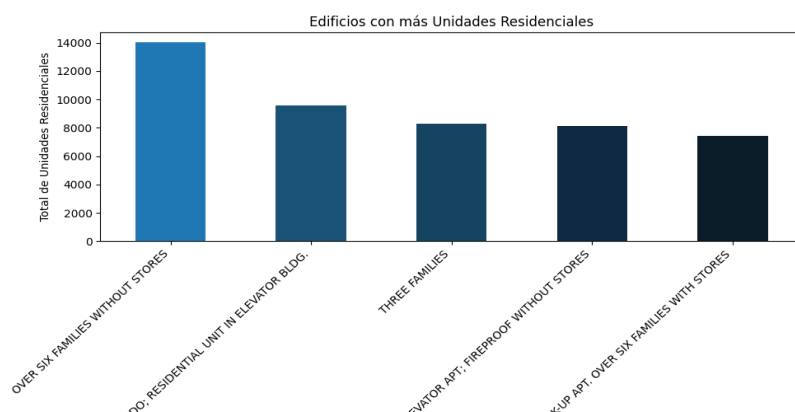
Se aplicó la prueba Kruskal-Wallis para determinar si existe alguna dependencia entre las variables, obteniendo un P-valor de 0. Hay suficiente evidencia estadística para asumir que hay una relación significativa entre el tipo de construcción y las unidades comerciales.

Ilustración 8. 5 edificios con más unidades comerciales



10.4. Unidades Residenciales contra Tipo de Construcción

Ilustración 9. 5 edificios con más unidades residenciales

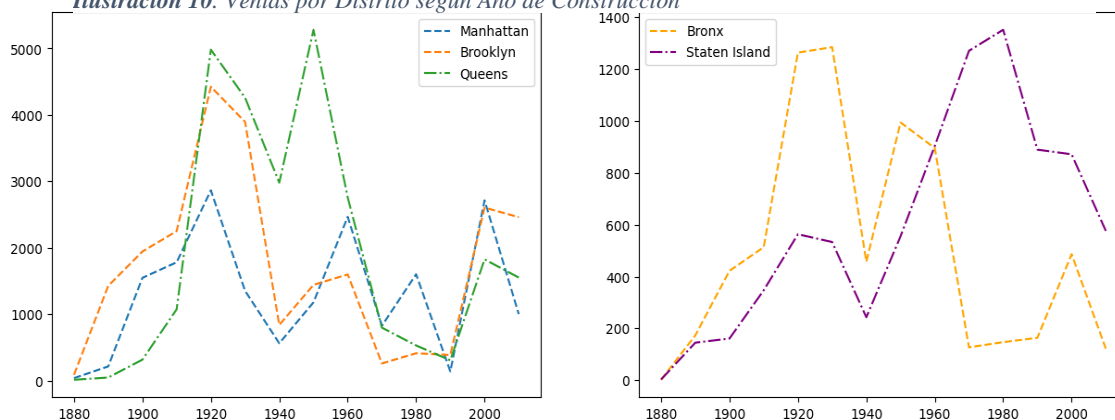


Se aplicó de igual forma la prueba de Kruskal-Wallis, la cual también arrojó un P-valor de 0, de igual manera, se infiere que se existe dependencia entre las unidades residenciales y el tipo de construcción.

10.5. Año de Construcción contra Distrito

En las gráficas mostradas⁸ se observa que los distritos mantienen la misma tendencia comercial según el año de construcción de la propiedad, mostrando los mismos picos y valles. De hecho, no se observan relaciones significativas entre las dos variables. Por Ejemplo: Queens posee la línea con mayor magnitud porque es el distrito con más propiedades registradas.

Ilustración 10. Ventas por Distrito según Año de Construcción

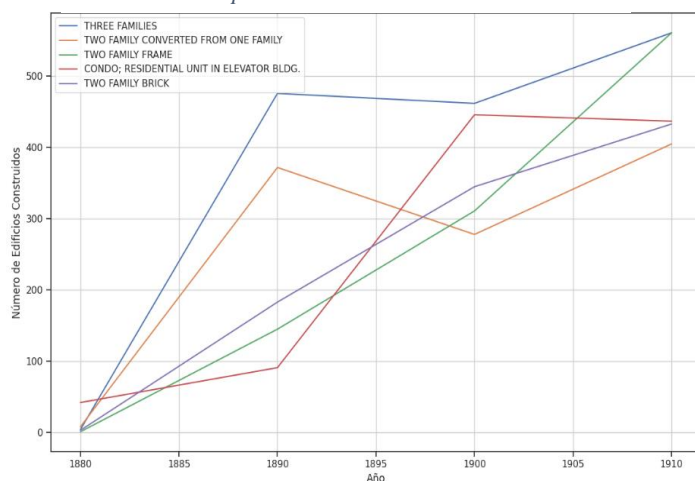


En esta medida, no se considera que exista alguna relación de dependencia entre ambas variables.

⁸ Nótese que cada gráfica está en escalas de 'y' diferentes

10.6. Año de Construcción contra Tipo de Construcción

Ilustración 11. Tipo de Construcción 1870 - 1910



1870-1910

Como se observa había aproximadamente 600 registros de los tipos de construcción en 40 años, sin embargo, con los datos que tenemos se observa que las viviendas de tres familias son las que más estaban siendo más construidas en la época.

1910-1950

Entre los años de 1910 a 1950 hay una mayor tendencia hacia los elevator cooperative, mientras que los *two family brick* y *two family converted from one family* que eran los edificios que reinaban en los años 1870 a 1910 han bajado significativamente en cuanto al número de construcciones.

Ilustración 12. Tipo de Construcción 1910 - 1950

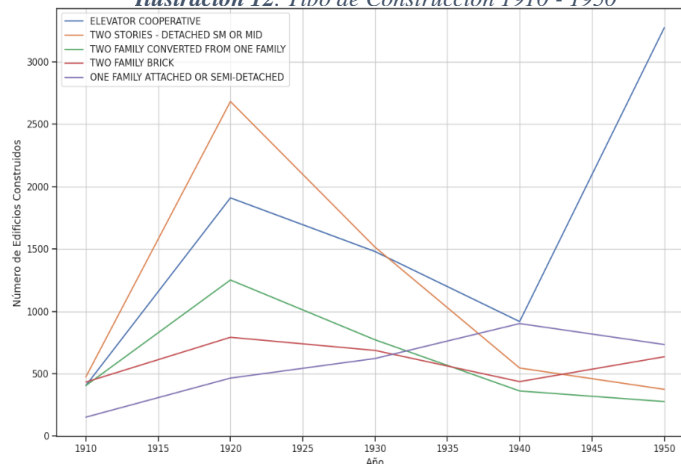
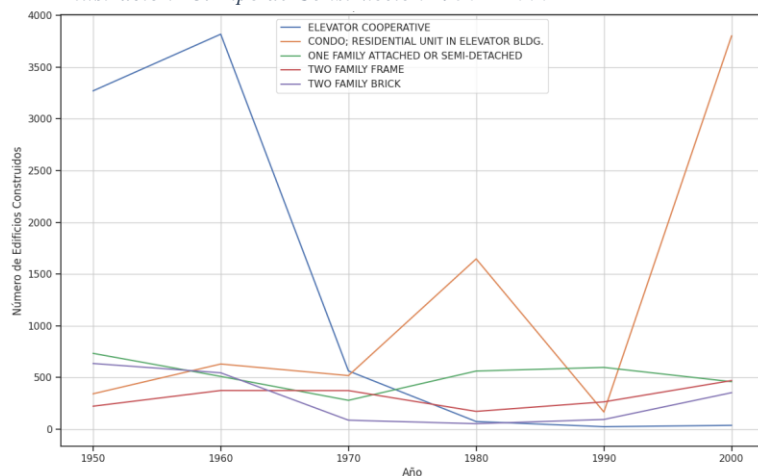


Ilustración 13. Tipo de Construcción 1950 - 2000

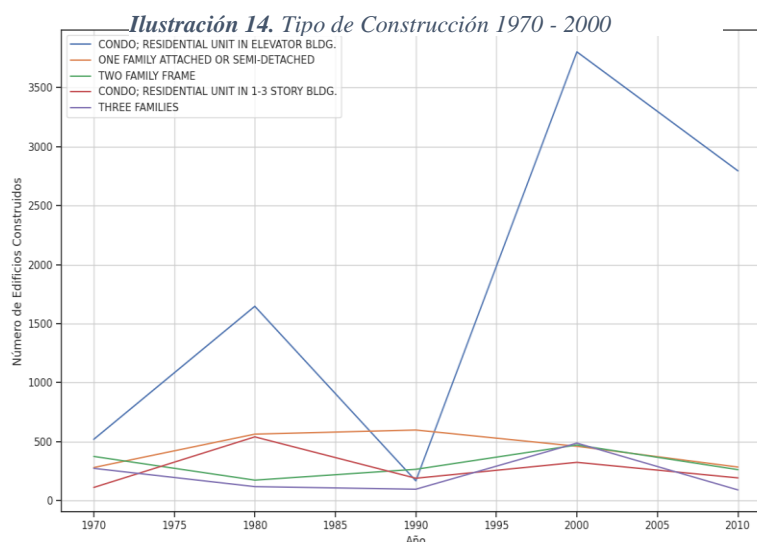


1950 - 2000

Se evidencia que desciende considerablemente los *elevator cooperative* teniendo en cuenta de que eran los que iban liderando en la época de los 50 's pero en los 60 's tienen su decrecimiento y se ponen en auge los *condos*.

1970 – 2010

En Ilustración 14 se puede evidenciar que las viviendas para tres familias volvieron a parecer y no había aparecido desde 1870/1910, de igual manera, los condominios siguen creciendo



En una investigación más profunda acerca del porqué en una época se construyen más edificios que otros y tratando de entender sus picos nos encontramos con varios factores:

- En 1900 la población neoyorquina era de 3'437.202⁹ y 10 años después, es decir, en 1910 Nueva York ya contaba con 9'113.614 habitantes¹⁰ esta constante expansión de la ciudad de Nueva York se debe a que en el período de 1910 a 1950 fue testigo de un rápido crecimiento urbano, avances tecnológicos en la construcción, cambios en la forma de vida de las personas que condujeron al auge de edificaciones como las viviendas cooperativas con ascensor.¹¹
- Factores como los cambios en las preferencias de vivienda, la evolución en el desarrollo urbano y demográfico puede haber contribuido al declive de las cooperativas con ascensor y el aumento de los condominios residenciales desde la década de 1950 hasta los años 2000

10.7. Año de Construcción contra Precio de Venta

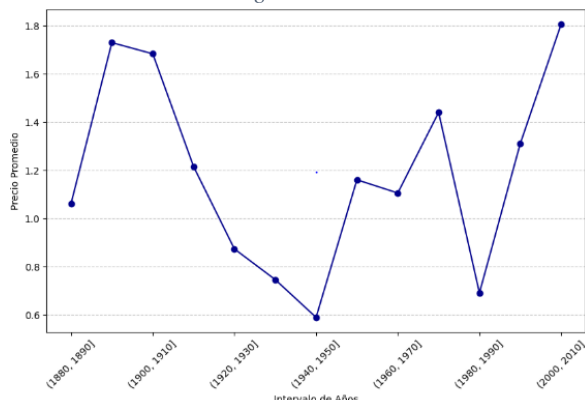
Se realizaron intervalos de tiempo por cada 10 años, y su respectivo promedio de precio de ventas registrados en 2016 y 2017.

⁹ (United States Census Bureau, s.f.)

¹⁰ (Datosmacro.com, s.f.)

¹¹ (MDZ Trip, 2021)

Ilustración 15. Precio según el Año de Construcción



En primer lugar, se observa un aumento significativo en los precios de venta de las propiedades construidas en la década de 1890. Este incremento podría estar relacionado con la ola masiva de inmigración que experimentó la ciudad de Nueva York en ese periodo¹². Con más de 8 millones de inmigrantes buscando hogar. Esto podría reflejar la valoración del mercado por estructuras con carácter y patrimonio histórico.

En segundo lugar, se evidencia un pico en los precios de venta alrededor de la década del 2000, particularmente en 2006. Este aumento puede atribuirse a la famosa burbuja inmobiliaria¹³ que afectó a más de la mitad de los estados de los Estados Unidos, lo que provocó un aumento desproporcionado en los precios de las viviendas.

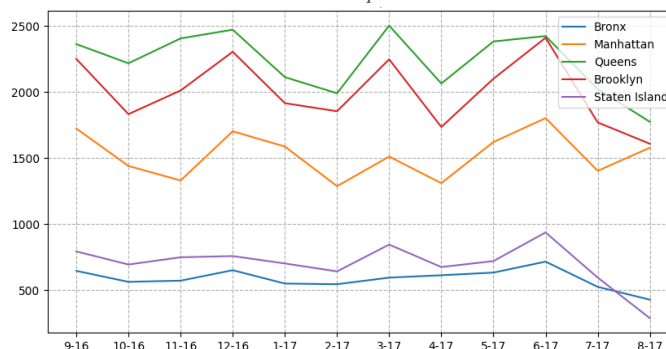
De igual manera en los años de 1900 a 1940 hay una fuerte caída de precios, una causa puede ser el pánico financiero de 1907, el cual se realizó cuando la Bolsa de Nueva York cayó casi 50% desde su máximo del año anterior¹⁴, esto puede afectar porque durante estos periodos de crisis financieras, los inversores consideran estas actividades y pueden retirar su capital de inversiones, como puede ser en el mercado inmobiliario, lo cual puede llevar a una disminución en la inversión de propiedades, de igual manera la obsolescencia de las estructuras en comparación con las construcciones más recientes baja los precios de las infraestructuras.

10.8. Distrito contra Fecha de Venta

Se agrupó la fecha por meses y se analizó la tendencia de ventas por distrito tal como se observa en la siguiente gráfica.

Todos los distritos mantienen las mismas tendencias en cuestión de picos y valles. Staten Island y Bronx mantienen un mercado menos volátil y más pausado que el observado en los tres distritos restantes; sin embargo, la tendencia en general es la misma. Centrándose en Queens, Manhattan y Brooklyn, se observan subidas en ventas importantes cada 3 meses (diciembre, marzo y junio) y caídas en febrero, abril y julio. Al parecer, las tendencias en el mercado están muy marcadas por épocas festivas como navidad, verano y pascua, que preceden caídas en ventas; tal como se evidenció en el análisis univariado en la variable de *Sale Price*.

Ilustración 16. Fecha de venta por Distrito



¹² (Sala, 2023)

¹³ (Telemundo, 2023)

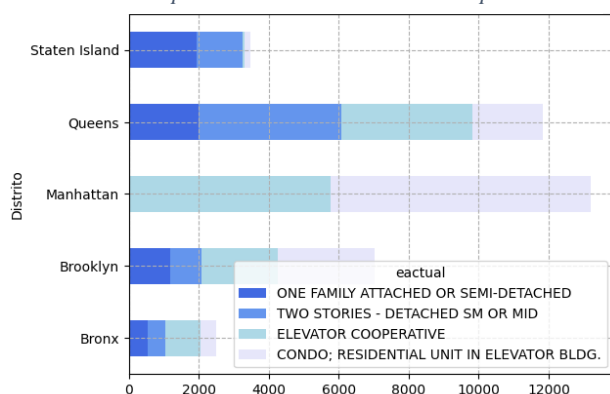
¹⁴ (FasterCapital, 2024)

10.9. Distrito contra Tipo de Construcción

Se aplicó la prueba chi – cuadrado de independencia para determinar la relación entre ambas variables, obteniendo un P valor de 0; es decir, se rechaza independencia y existe alguna relación entre ellas.

Para determinar la naturaleza de la dependencia, se identificaron los tipos de viviendas más comunes por distrito.

Ilustración 17. Tipo de Construcción más común por Distrito



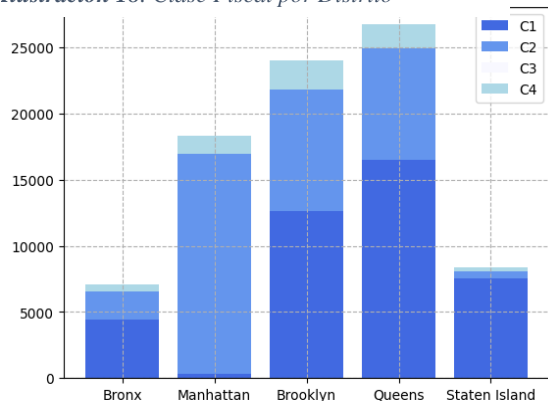
En Staten Island hay bastante preferencia por casas y residencias horizontales (para una familia o dos), por el contrario, Manhattan, Brooklyn y Bronx hay mayor cantidad de viviendas verticales (apartamentos y condominios con ascensor). Queens parece tener un equilibrio entre ambos tipos de residencias. El contraste previamente expresado se debe a que Staten Island suele ser el hogar de familias que buscan tranquilidad lejos de la urbe y el ritmo de vida en distritos más urbanizados como Manhattan

o Brooklyn.¹⁵

10.10. Distrito contra Clase Fiscal

Se aplicó la prueba chi – cuadrado de independencia para determinar la relación entre ambas variables, obteniendo un p valor de 0; es decir, existe alguna relación entre ellas.

Ilustración 18. Clase Fiscal por Distrito



Se observa que, en Manhattan, a diferencia del resto de distritos, son mayoritarios los edificios que pagan impuesto de tipo II (propiedades residenciales con más de 4 unidades) mientras que en el resto de los distritos predominan aquellos de tipo I (propiedades con menos de 4 unidades residenciales). Esta dependencia muestra cómo en Manhattan hay especial preferencia por propiedades con capacidad de albergar gran cantidad de unidades¹⁶ frente a otros distritos donde hay mayor atracción por propiedades para pocas

familias¹⁷. Esta relación está respaldada con otras dependencias como Distrito contra Tipo de Construcción.

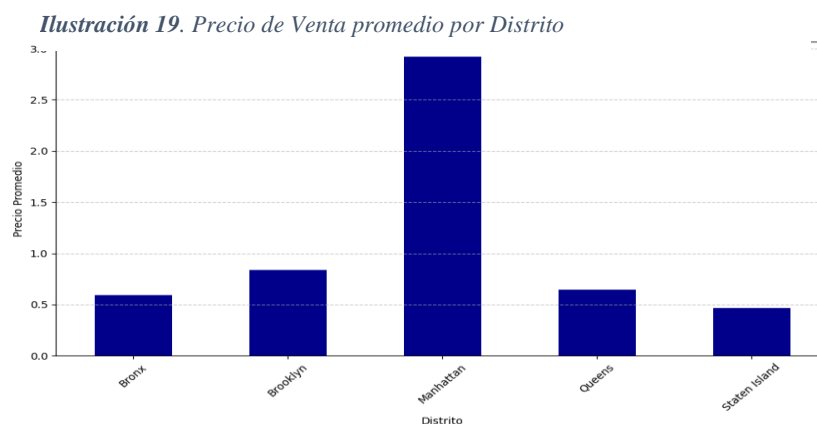
¹⁵ (Impacto Latino, 2018)

¹⁶ Como lo son la vivienda vertical

¹⁷ Como la vivienda horizontal

10.11. Distrito contra Precio de Venta

Se realiza la prueba Kruskal-Wallis para ver si hay dependencia entre las variables y con un P valor de 0 se rechaza H_0 , es decir, existe alguna relación entre el distrito y el precio de venta.



Se observa que Manhattan, siendo el distrito más pequeño, es el más poblado y donde se encuentran los principales centros comerciales, culturales y financieros. Además, posee los precios promedio de propiedades inmobiliarias más altos. Esto se debe a la gran demanda y al atractivo general de este distrito, que lo convierten en uno de los lugares más codiciados para vivir o invertir. Por otro lado, distritos como Staten Island y Bronx, que están más alejados del centro y tienen una menor densidad de población, tienden a tener precios más bajos en comparación con los distritos centrales¹⁸.

10.12. Clase fiscal contra Precio de Venta

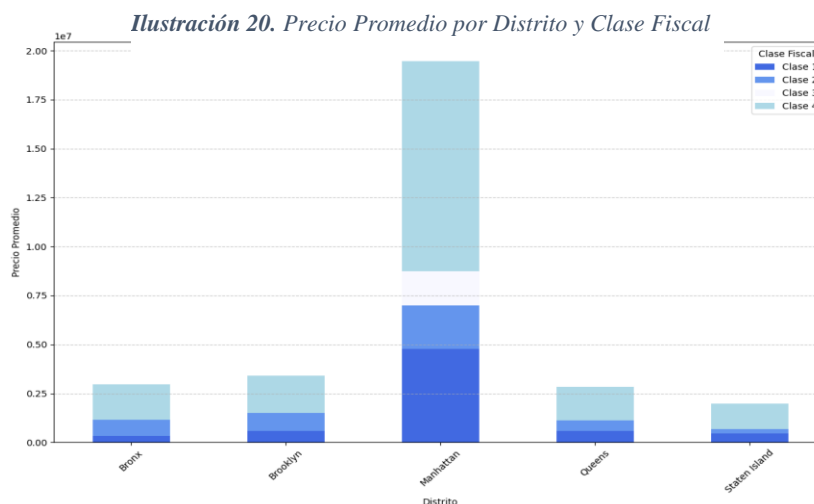
Se realizó la prueba Kruskal-Wallis para ver la dependencia entre las variables y con un P-valor de 0 hay relación entre la clase fiscal y el precio de venta.

Tabla 6. Precio Promedio de Venta por Clase Fiscal

<i>Clase Fiscal</i>	<i>Precio Promedio de Venta</i>
Clase 1	\$ 564,766.5
Clase 2	\$ 1'411,224
Clase 3	\$ 867,500
Clase 4	\$ 3'731,771

¹⁸ (Catalán, s.f.)

La tendencia señala que las propiedades comerciales y de mayor escala tienden a tener precios más altos que las residencias unifamiliares y pequeños comercios. Esto puede atribuirse a factores como la ubicación estratégica, las amenidades, el diseño y la calidad de construcción, características que suelen ser más exclusivas y orientadas a un mercado de mayor poder adquisitivo. La Ilustración 20 brinda visualización clara de cómo ciertas clases se concentran en distritos específicos.

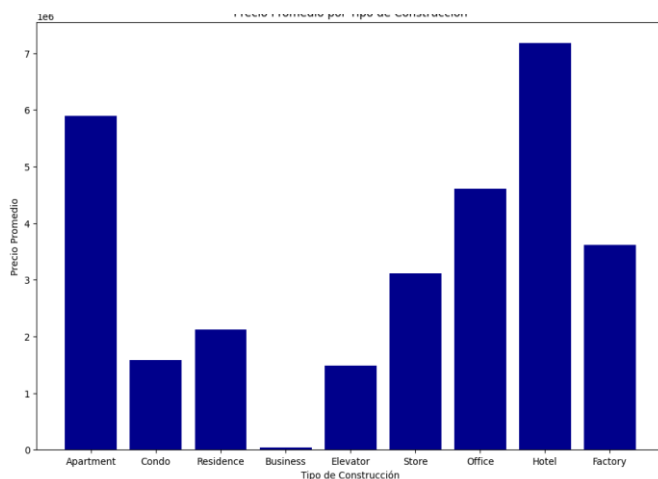


10.13. Tipo de construcción contra Precio de Venta.

Se realizó la prueba Kruskal-Wallis para ver si son dependientes o independientes y con un P-valor de 0 el tipo de construcción está relacionado con el precio de venta

Dado los datos entregados, hay demasiadas categorías del tipo de construcción, por este motivo se hicieron subgrupos de ciertas características para estudiarlas de manera más detallada.

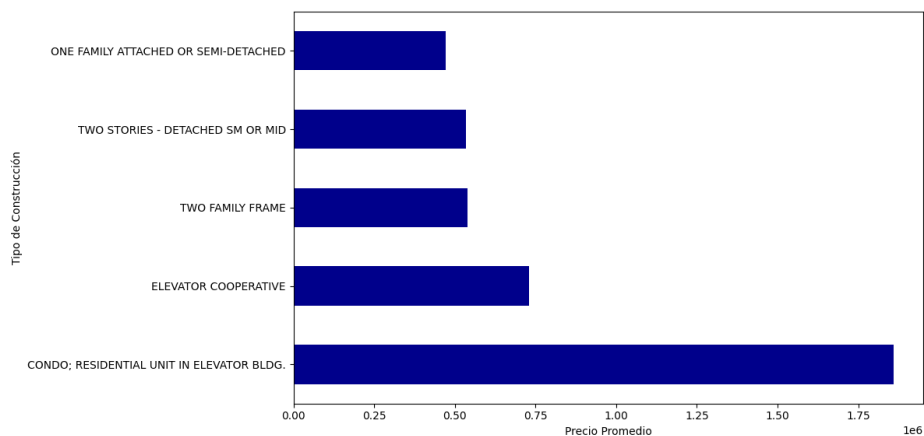
Ilustración 21. Precio promedio por Tipo de Construcción



Ciertas categorías de construcción, como hoteles y condominios, tienden a tener precios de venta más altos en comparación con otras tipologías, como apartamentos y casas unifamiliares. Esto puede deberse a factores como la ubicación y el diseño de estas propiedades, que suelen ser más exclusivas y orientadas a un mercado de mayor poder adquisitivo.

Para observar de otra manera el tipo de construcción se tomaron las cinco características más nombradas en los datos:

Ilustración 22. Precio Promedio de las 5 edificaciones más comunes



Al ser condominios o unidades con mayores características contemplan un costo mayor a diferencia de otras construcciones más sencillas o amplias, de igual manera se ve afectado el promedio de precio de ventas dependiendo su localidad.

11. Resumen de las pruebas del Análisis Bivariado

Tabla 7. Resumen - Pruebas de Dependencia y Correlación¹⁹

Variables	Distrito	Clase Fiscal	Tipo de Construcción	Unidades Residenciales	Unidades Comerciales	Precio de Venta
Distrito	D	D	D	D	D	D
Clase Fiscal	D	D	D	D	D	D
Tipo de Construcción	D	D	D	D	D	D
Unidades Residenciales	D	D	D	1	0.035 Spearman 0.032 Kendall	0.048 Spearman 0.041 Kendall
Unidades Comerciales	D	D	D	0.035 Spearman 0.032 Kendall	1	0.026 Spearman 0.02 Kendall
Precio de Venta	D	D	D	0.048 Spearman 0.041 Kendall	0.026 Spearman 0.02 Kendall	1

¹⁹ D: Dependientes

I: Independientes

12. Conclusiones

Una vez realizado el análisis exploratorio de los datos, se concluye:

- El distrito de Manhattan tiene los precios más elevados debido a que es un punto central altamente demandado por el sector financiero, turístico e inmobiliario. Además, es el que posee más unidades comerciales y es la zona en la que más se pagan impuestos a precios elevados.
- A partir de 1990 y hasta 2010, los condominios de apartamentos son el tipo de edificación más construida y demandada en los distritos centrales como Manhattan y Brooklyn; en cambio, en Staten Island, caracterizado por ser una zona más residencial con viviendas horizontales y unifamiliares, casi no se encuentran este tipo de edificaciones. En general, áreas con alto grado de desarrollo urbano y comercial tienden a preferir la vivienda vertical (ya que almacena mayor número de unidades residenciales).
- Los hoteles encabezan los precios de venta, situación que está respaldada no solo por la infraestructura sino por la locación y alta demanda comercial a la que está expuesto. Después de los hoteles, los apartamentos son los más costosos debido a la amplia variedad de precios (desde residencias de lujo hasta compartidas).
- Los picos de ventas de las edificaciones se encuentran en los meses de diciembre, junio y marzo, ya que ofrecen a los compradores mejores ofertas, lo que motiva la compra de inmuebles. Además, estas temporadas hay mayor disposición y facilidad en transacciones comerciales por corresponder a la primavera y navidad.
- Se puede encontrar que mayormente la edificaciones de clase fiscal 1 y 2 las cuales se caracterizan por ser residenciales; sin embargo, dependiendo el distrito donde se encuentre este aspecto puede variar, resaltando que Manhattan y Brooklyn son más comerciales a diferencia de otros distritos que se encuentran en Nueva York.

En general, Nueva York presenta una amplia oferta inmobiliaria y la inversión más rentable depende fundamentalmente de los intereses del individuo; por ejemplo, si se busca una vivienda familiar, la mejor oportunidad está en Staten Island, donde se garantizan bajos precios a cambio de una locación tranquila y lejana al estrés urbano; por otro lado, si se prefiere un inversión con fines meramente monetarios, Manhattan o Brooklyn son la mejor alternativa ya que la rentabilidad está prácticamente asegurada.

13. Bibliografía

- Catalán, I. (s.f.). *Los 6 mejores barrios de Nueva York*. Obtenido de <https://www.hellotickets.com.co/estados-unidos/nueva-york/barrios/sc-1-1564>
- Cifuentes, C. (5 de December de 2022). *La estrategia de Nueva York para sacar a las personas sin hogar de las calles*. Recuperado el 13 de May de 2024, de La Tercera: <https://www.latercera.com/mundo/noticia/la-estrategia-de-nueva-york-para-sacar-a-las-personas-sin-hogar-de-las-calles/RQ3EY3RMKNE73DIYJYFEVWNTA/>
- Datosmacro.com. (s.f.). *Población de Nueva York*. Obtenido de <https://datosmacro.expansion.com/demografia/poblacion/usa-estados/nueva-york>
- FasterCapital. (18 de Marzo de 2024). *Desentranando las causas y consecuencias del panico del banco de 1907*. Obtenido de <https://fastercapital.com/es/contenido/Desentranando-las-causas-y-consecuencias-del-panico-del-banco-de-1907.html#factores-econ-micos-y-pol-ticos>
- Guiajando. (s.f.). *Flushing, el verdadero Chinatown de NY en Queens*. Recuperado el 13 de May de 2024, de Newyorkando: <https://www.newyorkando.com/flushing/>
- Impacto Latino. (24 de February de 2018). *Los mejores barrios para vivir en Staten Island*. Recuperado el 13 de May de 2024, de Impacto Latino: <https://impactolatino.com/los-mejores-barrios-vivir-staten-island/>
- Jordan, M., & Tyson, M. (s.f.). *Bed-Stuy, Brooklyn*. Recuperado el 13 de May de 2024, de Newyorkando: <https://www.newyorkando.com/bedford-stuyvesant/>
- MDZ Trip. (27 de Julio de 2021). *Cómo era la mágica ciudad de Nueva York a principios del siglo XX, ¡imágenes sorprendentes!* Obtenido de <https://www.mdzol.com/estilo/2021/7/27/como-era-la-magica-ciudad-de-nueva-york-principios-del-siglo-xx-imagenes-sorprendentes-175418.html>
- Nuevayork.es. (s.f.). *Upper East Side en Nueva York Upper East Side en Nueva York*. Recuperado el 13 de May de 2024, de NuevaYork.es: <https://www.nuevayork.es/practico/barrios/upper-east-side/>
- NYC. (s.f.). *Glossary of Terms for Property Sales Files*. Obtenido de <https://www.nyc.gov/site/finance/property/glossary-property-sales.page>
- NYC. (s.f.). *glossary-property-sales*. Recuperado el 14 de May de 2024, de NYC.gov: <https://www.nyc.gov/site/finance/property/glossary-property-sales.page>
- Sala, A. (23 de Febrero de 2023). *Desembarco de inmigrantes en ellis island, nueva york*. Obtenido de https://historia.nationalgeographic.com.es/edicion-impres/a/articulos/desembarco-de-inmigrantes-en-ellis-island-nueva-york_19031#google_vignette
- Telemundo. (24 de Julio de 2023). *El hombre que anticipó la burbuja inmobiliaria de 2008 dice qué pasará con los precios de la vivienda en EE.UU.* Obtenido de <https://www.telemundo.com/noticias/noticias-telemundo/economia/robert-shiller-el-hombre-que-antipico-la-burbuja-inmobiliaria-de-2008-d-rcna96053>
- United States Census Bureau. (s.f.). *POP Culture: 1900*. Obtenido de https://www.census.gov/history/www/through_the_decades/fast_facts/1900_fast_facts.html

Universidad Nacional de Cuyo. (s.f.). *Analítica de Datos 4.0*. Recuperado el 14 de May de 2024, de Wikipedia: https://themys.sid.uncu.edu.ar/rpalma/Logistica/Analitica_40/index.html