

HDB Resale Prices in Singapore EDA

HDB resale prices in Singapore have been on a steady incline for the past few years, with no indication of this trend changing in the near future. This has caused some concern among home buyers, as the high prices may prevent them from being able to afford a home in the city-state. Despite this, demand for HDB resale units continues to be high, as many people view property ownership as a sound investment.

This paper seeks to explore:

- The reasons behind this trend
- If the trend is likely to continue
- Whether a purchase of a 4-room Model A flat that is 25 years and above will still have its resale value appreciated/depreciated

Official data of the resale HDB units were sourced from: <https://data.gov.sg/dataset/resale-flat-prices>

```
df_2012 = pd.read_csv(r"C:\Users\benyo\Dropbox\PC\Downloads\ResaleHDBPrices\resale-flat-prices-based-on-approval-date-2008-feb-2012.csv")
df_2014 = pd.read_csv(r"C:\Users\benyo\Dropbox\PC\Downloads\ResaleHDBPrices\resale-flat-prices-based-on-registration-date-from-mar-2012-to-dec-2014.csv")
df_2016 = pd.read_csv(r"C:\Users\benyo\Dropbox\PC\Downloads\ResaleHDBPrices\resale-flat-prices-based-on-registration-date-from-jan-2015-to-dec-2016.csv")
df_2017 = pd.read_csv(r"C:\Users\benyo\Dropbox\PC\Downloads\ResaleHDBPrices\resale-flat-prices-based-on-registration-date-from-jan-2017-onwards.csv")

consolidated_resale_data = pd.concat([df_2012, df_2014, df_2016, df_2017])
consolidated_resale_data.shape

(584471, 11)

In [3]: consolidated_resale_data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 584471 entries, 0 to 125463
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  --
0   month               584471 non-null    object
1   town               584471 non-null    object
2   flat_type          584471 non-null    object
3   block              584471 non-null    object
4   street_name        584471 non-null    object
5   storey_range       584471 non-null    object
6   floor_area_sqm     584471 non-null    float64
7   flat_model         584471 non-null    object
8   lease_commence_date 584471 non-null    int64
9   resale_price       584471 non-null    float64
10  remaining_lease    162617 non-null    object
dtypes: float64(2), int64(1), object(8)
memory usage: 53.5+ MB

In [4]: consolidated_resale_data.head(10)

Out[4]:
```

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price	remaining_lease
0	2000-01	ANG MO KIO	3 ROOM	170	ANG MO KIO AVE 4	07 TO 09	69.0	Improved	1986	147000.0	NaN
1	2000-01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	04 TO 06	61.0	Improved	1986	144000.0	NaN
2	2000-01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	159000.0	NaN
3	2000-01	ANG MO KIO	3 ROOM	215	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	167000.0	NaN
4	2000-01	ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	07 TO 09	67.0	New Generation	1976	163000.0	NaN
5	2000-01	ANG MO KIO	3 ROOM	320	ANG MO KIO AVE 1	04 TO 06	73.0	New Generation	1977	157000.0	NaN
6	2000-01	ANG MO KIO	3 ROOM	320	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1977	178000.0	NaN
7	2000-01	ANG MO KIO	3 ROOM	330	ANG MO KIO AVE 1	07 TO 09	68.0	New Generation	1981	160000.0	NaN
8	2000-01	ANG MO KIO	3 ROOM	330	ANG MO KIO AVE 1	04 TO 06	68.0	New Generation	1981	169000.0	NaN
9	2000-01	ANG MO KIO	3 ROOM	332	ANG MO KIO AVE 1	07 TO 09	82.0	New Generation	1981	205000.0	NaN

We can see from the info from the previous codes that a big portion of data on remaining lease of the sold resale units are missing, most likely from the older data records that might not have a need for such data at that point in time. To have a more accurate analysis, we will fill in the missing data by subtracting the lease commencement date from the transaction year

```
In [5]: ## Extract the year from the column "month".

consolidated_resale_data[["transaction_year", "month"]] = consolidated_resale_data.month.str.split(pat='-', expand = True)
consolidated_resale_data.head()
```

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price	remaining_lease	transaction_year
0	01	ANG MO KIO	3 ROOM	170	ANG MO KIO AVE 4	07 TO 09	69.0	Improved	1986	147000.0	NaN	2000
1	01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	04 TO 06	61.0	Improved	1986	144000.0	NaN	2000
2	01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	159000.0	NaN	2000
3	01	ANG MO KIO	3 ROOM	215	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	167000.0	NaN	2000
4	01	ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	07 TO 09	67.0	New Generation	1976	163000.0	NaN	2000

```
In [6]: ## Convert "year" value into integer.

consolidated_resale_data[["transaction_year"]] = pd.to_numeric(consolidated_resale_data[["transaction_year"]])
consolidated_resale_data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 584471 entries, 0 to 125463
Data columns (total 12 columns):
```

We can see from the info from the previous codes that a big portion of data on remaining lease of the sold resale units are missing, most likely from the older data records that might not have a need for such data at that point in time. To have a more accurate analysis, we will fill in the missing data by subtracting the lease commencement date from the transaction year

```
6 floor_area_sqm      584471 non-null float64
7 flat_model          584471 non-null object
8 lease_commence_date 584471 non-null int64
9 resale_price        584471 non-null float64
10 remaining_lease     162617 non-null object
11 transaction_year    584471 non-null int64
dtypes: float64(2), int64(2), object(8)
memory usage: 58.0+ MB
```

```
In [7]: df = pd.DataFrame(consolidated_resale_data)
df.head()
```

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price	remaining_lease	transaction_year
0	01	ANG MO KIO	3 ROOM	170	ANG MO KIO AVE 4	07 TO 09	69.0	Improved	1986	147000.0	NaN	2000
1	01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	04 TO 06	61.0	Improved	1986	144000.0	NaN	2000
2	01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	159000.0	NaN	2000
3	01	ANG MO KIO	3 ROOM	215	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	167000.0	NaN	2000
4	01	ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	07 TO 09	67.0	New Generation	1976	163000.0	NaN	2000

We obtain the number of years remaining on the lease on the flats by subtracting the age of the flats when they were sold from the maximum 99years lease given by HDB.

```
In [8]: df['remaining_lease_when_sold'] = 99 - (df['transaction_year'] - df['lease_commence_date'])
df.head()
```

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price	remaining_lease	transaction_year	remaining_lease_when_sold
--	-------	------	-----------	-------	-------------	--------------	----------------	------------	---------------------	--------------	-----------------	------------------	---------------------------

We obtain the number of years remaining on the lease on the flats by subtracting the age of the flats when they were sold from the maximum 99years lease given by HDB.

1	01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	04 TO 06	61.0	Improved	1986	144000.0	NaN	2000	85
2	01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	159000.0	NaN	2000	75
3	01	ANG MO KIO	3 ROOM	215	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	167000.0	NaN	2000	75
4	01	ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	07 TO 09	67.0	New Generation	1976	163000.0	NaN	2000	75

```
In [9]: ## Drop the 'month' and 'remaining_lease' columns as they are now irrelevant. Drop 'block' as well.
df = df.drop(columns=['month','remaining_lease','lease_commence_date','block'])
print(df)
```

	town	flat_type	street_name	storey_range \
0	ANG MO KIO	3 ROOM	ANG MO KIO AVE 4	07 TO 09
1	ANG MO KIO	3 ROOM	ANG MO KIO AVE 4	04 TO 06
2	ANG MO KIO	3 ROOM	ANG MO KIO AVE 1	07 TO 09
3	ANG MO KIO	3 ROOM	ANG MO KIO AVE 1	07 TO 09
4	ANG MO KIO	3 ROOM	ANG MO KIO AVE 1	07 TO 09
...
125459	YISHUN	EXECUTIVE	YISHUN ST 81	01 TO 03
125460	YISHUN	EXECUTIVE	YISHUN ST 81	07 TO 09
125461	YISHUN	MULTI-GENERATION	YISHUN ST 61	04 TO 06
125462	YISHUN	MULTI-GENERATION	YISHUN ST 61	04 TO 06
125463	YISHUN	MULTI-GENERATION	YISHUN ST 61	16 TO 12

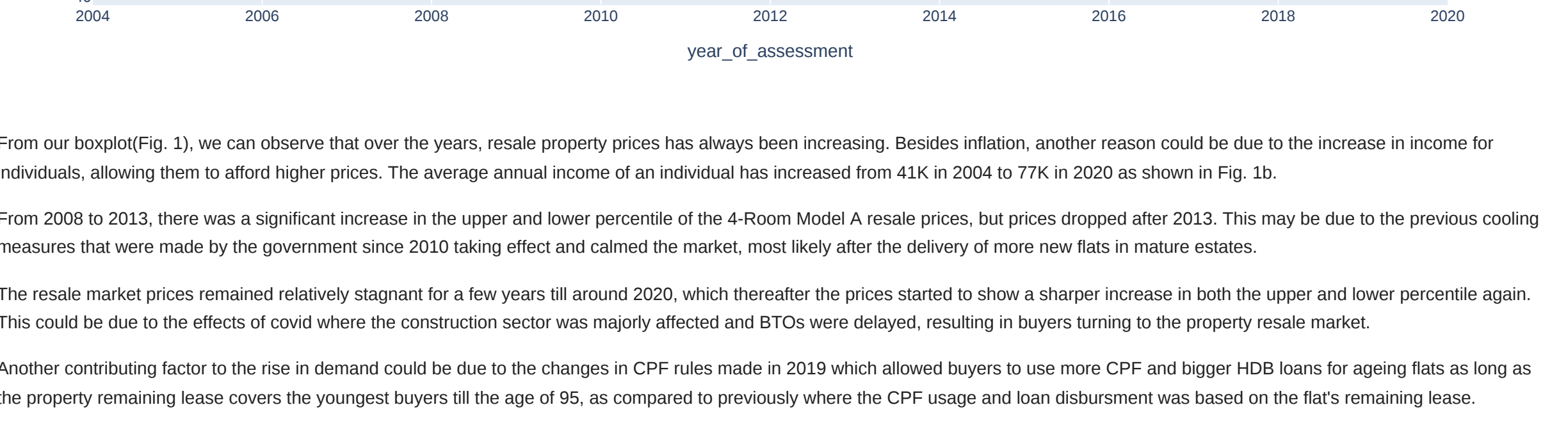
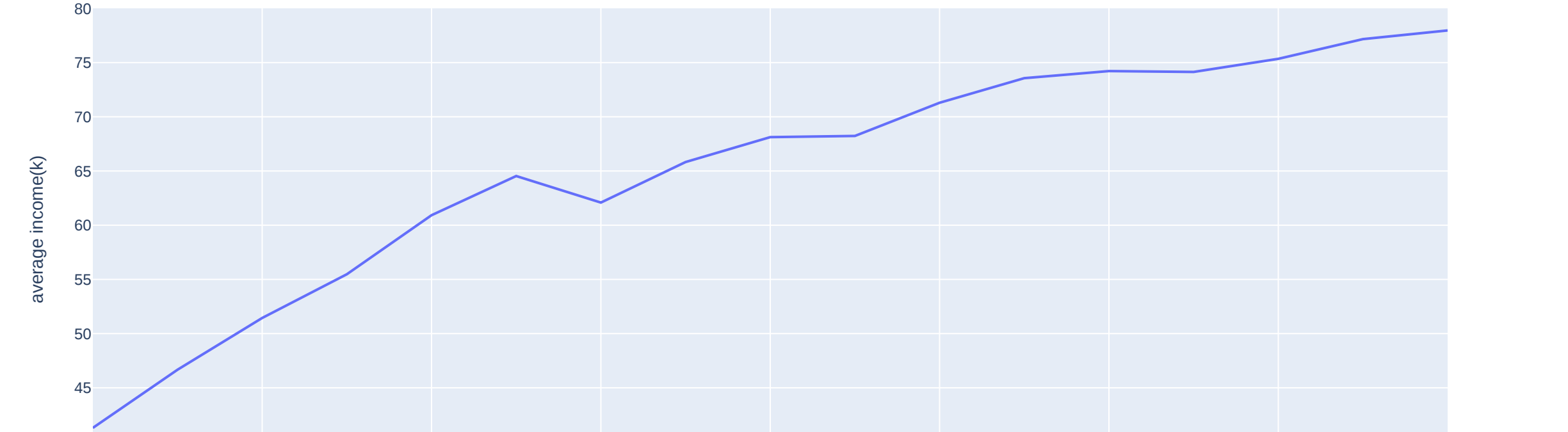
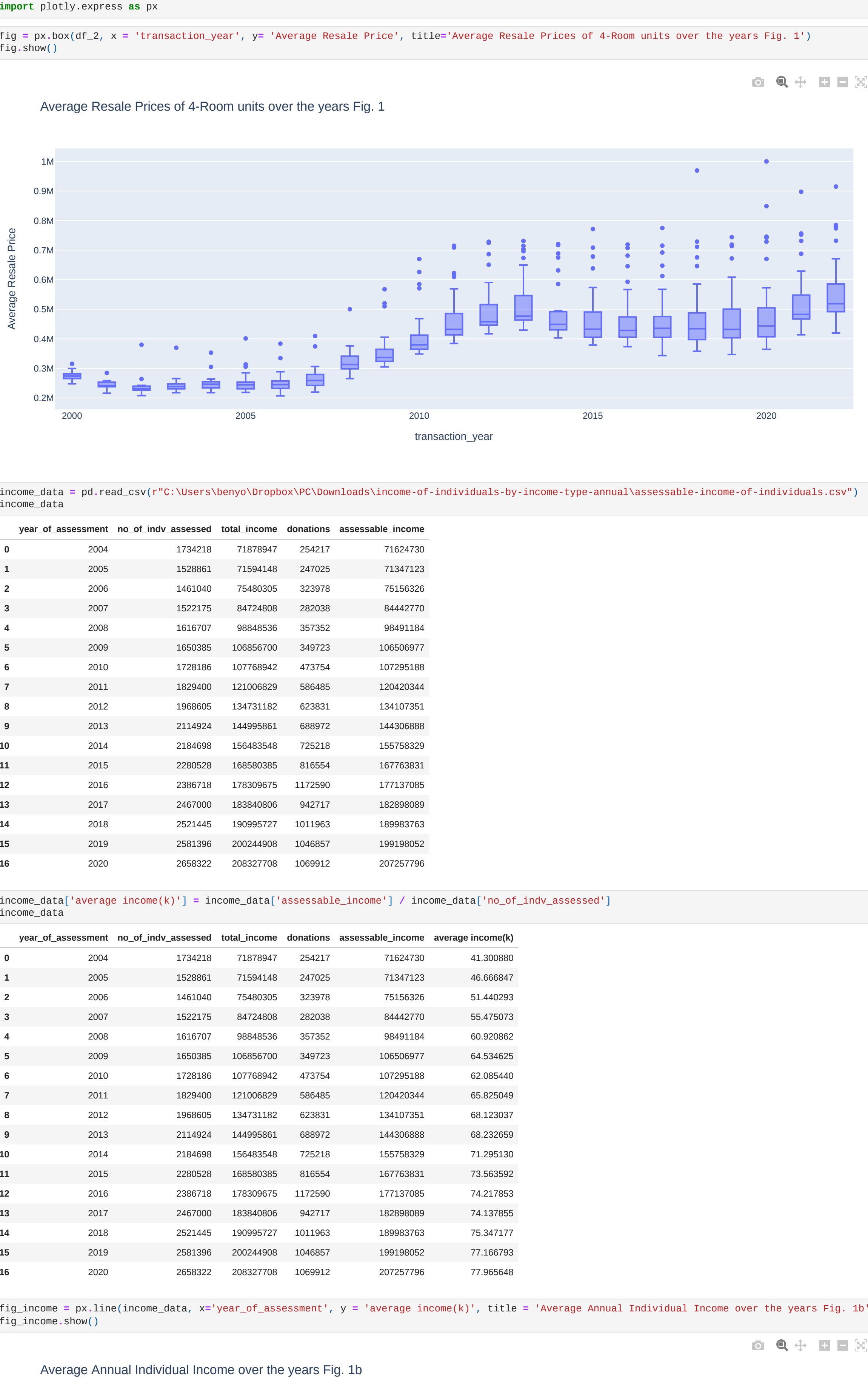
```
0      floor_area_sqm      flat_model      resale_price      transaction_year \
0      69.0      Improved      147899.0      2989
1      61.0      Improved      144899.0      2921
2      73.0      New Generation      159899.0      2989
3      73.0      New Generation      167899.0      2989
4      67.0      New Generation      163899.0      2989
...      ...      ...      ...      ...
125459      142.0      Apartment      739389.0      2922
125460      142.0      Apartment      773989.0      2922
125461      164.0      Multi Generation      785989.0      2922
125462      171.0      Multi Generation      842989.0      2922
125463      164.0      Multi Generation      845989.0      2922
...      ...      ...      ...      ...
0      85      85      85      85
1      75      75      75      75
2      75      75      75      75
3      75      75      75      75
4      75      75      75      75
...      ...      ...      ...      ...
125459      65      65      65      65
125460      65      65      65      65
125461      64      64      64      64
125462      64      64      64      64
125463      64      64      64      64
...      ...      ...      ...      ...
```

```
In [10]: [584471 rows x 9 columns]
df_grouped = df.groupby(["flat_type","flat_model","remaining_lease_when_sold","transaction_year"])[["resale_price"]].mean()
df_grouped = pd.DataFrame(df_grouped)
print(df_grouped)
```

	flat_type	flat_model	remaining_lease_when_sold	transaction_year	resale_price
0	1 ROOM	Improved	52	2022	213000.000000
1			53	2021	188999.090999
2			54	2020	197222.222222
3			55	2019	180000.000000
4			56	2018	178842.545455
5			56	2018	184111.111111
...
7739	MULTI-GENERATION	Multi Generation	83	2003	431333.333333
7740			84	2002	431885.714286
7741			85	2001	468664.705882
7742			86	2000	500050.000000
7743			86	2001	431166.666667

```
In [11]: [7744 rows x 5 columns]
df_grouped.rename(columns={"resale_price": "Average Resale Price"}, inplace = True)
df_grouped = df_grouped.reset_index()
```

	flat_type	flat_model	remaining_lease_when_sold	transaction_year	Average Resale Price
0	1 ROOM	Improved	52	2022	213000.000000
1			53	2021	197222.222222
2			54	2020	180000.000000
3			55	2019	178842.545455
4			56	2018	184111.111111
...
7739	MULTI-GENERATION	Multi Generation	83	2003	431333.333333
7740			84	2002	431885.714286
7741			85	2001	468664.705882
7742			86	2000	500050.000000
7743			86	2001	431166.666667



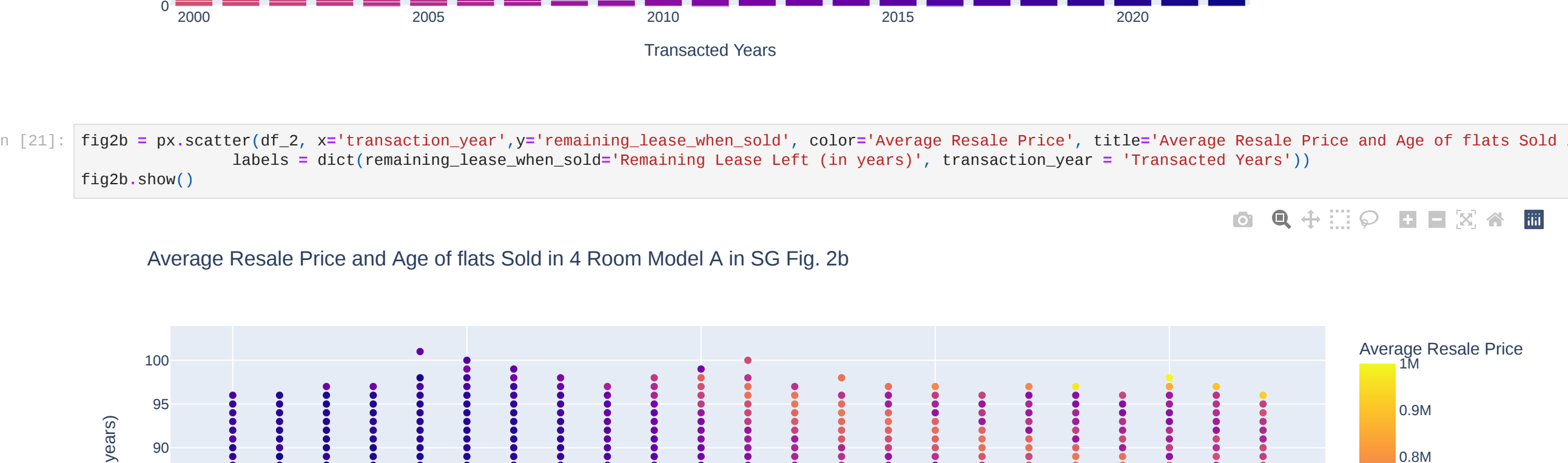
From our boxplot (Fig. 1), we can observe that over the years, resale property prices has always been increasing. Besides inflation, another reason could be due to the increase in income for individuals, allowing them to afford higher prices. The average annual income of an individual has increased from 41K in 2004 to 77K in 2020 as shown in Fig. 1b.

From 2008 to 2013, there was a significant increase in the upper and lower percentile of the 4-Room Model A resale prices, but prices dropped after 2013. This may be due to the previous cooling measures that were made by the government since 2010 taking effect and calmed the market, most likely after the delivery of more new flats in mature estates.

The resale market prices remained relatively stagnant for a few years till around 2013, where thereafter the prices started to show a sharper increase in both the upper and lower percentile again. This could be due to the effects of covid where the construction sector was majorly affected and BTOs were delayed, resulting in buyers turning to the property resale market.

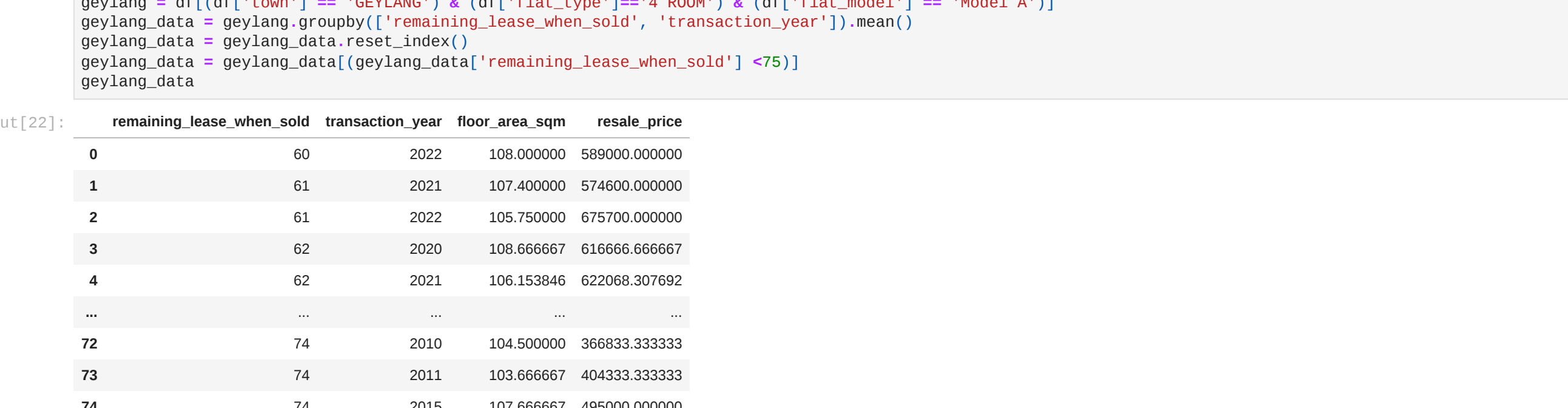
Another contributing factor to the rise in demand could be due to the changes in CPF rules made in 2019 which allowed buyers to use more CPF and bigger HDB loans for ageing flats as long as the property remaining lease covers the youngest buyers till the age of 95, as compared to previously where the CPF usage and loan disbursement was based on the flats' remaining lease.

The government has started to introduce new measures in second half of 2021 to cool the market again, however this uptrend in the property prices may likely continue as it also took a few years for the previous property measures taken in 2010 to have a visible effect. It might be possible to see the market stabilize in 3 to 4 years, around 2024/25, similarly to the previous property boom pattern.



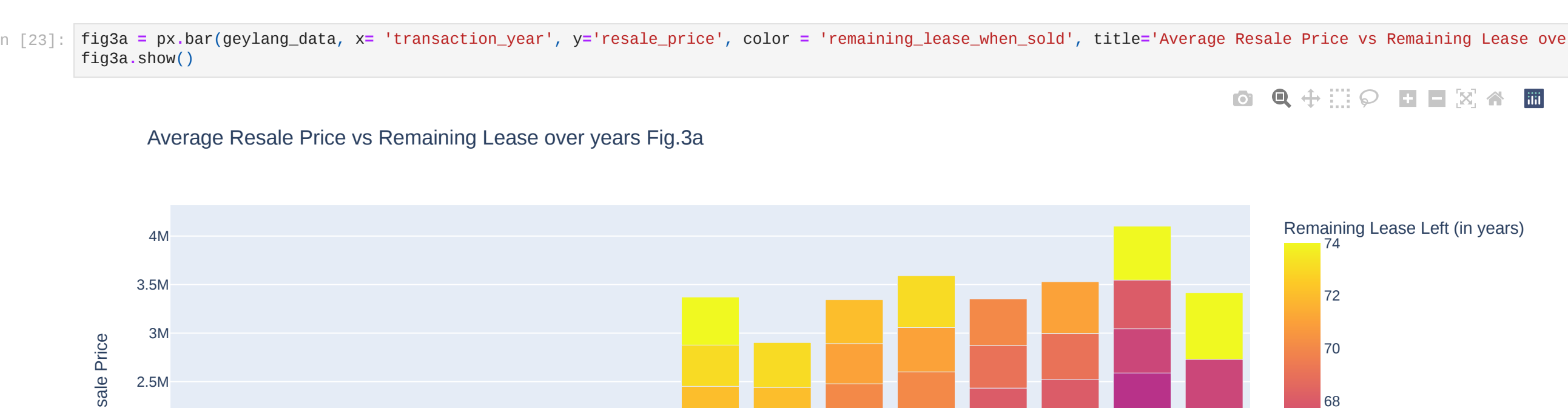
From the figures 2a & 2b above, it seems that on average older flats are still getting sold and their resale prices still appreciates over the years.

Let's dive into a mature estate sample, Geylang, and check whether a flat that is 24 years old and above will have its resale value appreciated, depreciated or hold its value.



From figure 3a, it seems that from 2021, there are more 24 years and older 4 room model A flats sold, with an increase in 2022 even when at this point where our data is collected in 2022.

Let's zoom in to 34year old and above flats.



For flats that are 25 year old and above in Geylang, there is a year on year increase of around 10% in their resale prices between 2021 and 2022.

If the current trends continue, 4-room flats that are 25 year old and above may still have their resale value appreciated.

