Boston Airbnb Detailed analysis of neighbourhoods

For details on how to run the file refer the Project Report, particularly the How to Run section

In this analysis, we will be looking at the three neighbourhoods, Fenway, Allston and Dorchester, which were considered as some of the poorest neighbourhoods in boston, and see how Airbnb listings have increased over. the years in each of thesew neighbourhoods, and thus inevitable contributing to the economy of the neighbourhood ¶

In [1]:
```
from pyspark.sql import SparkSession
from pyspark.sql import functions as F
from pyspark.sql.types import *
import pandas as pd
from pyspark.sql import types as T
from pyspark import SparkFiles
from textblob import TextBlob
```

In [2]:
```
spark = SparkSession\
        .builder\
        .master("local")\
        .config("spark.files.overwrite", True)\
        .getOrCreate()
```

In [3]:
```
def get_complete_boston_data(listings_url, reviews_url, year):
    boston_reviews_df = spark.read.option("header", "true").csv(reviews_url
    boston_listings_df = spark.read.csv(listings_url, header=True)
    boston_reviews_df = boston_reviews_df.where("comments != 'None'")
    boston_reviews_df = boston_reviews_df.withColumnRenamed("id", "review_i
    boston_top_neighbourhood = boston_listings_df.join(boston_neighbourhood
    boston_top_neighbourhood = boston_top_neighbourhood.join(boston_reviews
    return boston_top_neighbourhood.withColumn("year", F.lit(year))
```

In [4]:
```
neighbourhoods_list = ['Fenway', 'Allston', 'Dorchester']
boston_neighbourhoods_df = spark.createDataFrame(neighbourhoods_list, Strin
boston_neighbourhoods_df = boston_neighbourhoods_df.selectExpr("value as ne
```

In [5]:
```
reviews_url = "data/Boston/2020/June/reviews.csv"
listings_url = "data/Boston/2020/June/listings_original.csv"

boston_neighbourhood_2020 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2020.toPandas()
```

Out[5]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	latitude
0	Dorchester	18711	The Dorset Redline | 3BR 1BA | Walk to Redline...	71783	Lance	None	42.32212
1	Dorchester	18711	The Dorset Redline | 3BR 1BA | Walk to Redline...	71783	Lance	None	42.32212
2	Dorchester	18711	The Dorset Redline | 3BR 1BA | Walk to Redline...	71783	Lance	None	42.32212
3	Dorchester	18711	The Dorset Redline | 3BR 1BA | Walk to Redline...	71783	Lance	None	42.32212
4	Dorchester	18711	The Dorset Redline | 3BR 1BA | Walk to Redline...	71783	Lance	None	42.32212
...	...	...	...	...	...	...	...
29085	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	42.32112
29086	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	42.32112
29087	Dorchester	42831086	Lovely Home | near T	161724560	David	None	42.31952

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	latitude
29088	Dorchester	42831086	Lovely Home | near T	161724560	David	None	42.31952
29089	Dorchester	42831086	Lovely Home | near T	161724560	David	None	42.31952

29090 rows × 23 columns

In [6]:
```
reviews_url = "data/Boston/2019/June/reviews.csv"
listings_url = "data/Boston/2019/June/listings.csv"
boston_neighbourhood_2019 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2019.toPandas()
```

Out[6]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	latitude
0	Dorchester	18711	The Dorset | 3BR 1BA | 1st Floor	71783	Lance	None	42.32329
1	Dorchester	18711	The Dorset | 3BR 1BA | 1st Floor	71783	Lance	None	42.32329
2	Dorchester	18711	The Dorset | 3BR 1BA | 1st Floor	71783	Lance	None	42.32329
3	Dorchester	18711	The Dorset | 3BR 1BA | 1st Floor	71783	Lance	None	42.32329
4	Dorchester	18711	The Dorset | 3BR 1BA | 1st Floor	71783	Lance	None	42.32329
...	...	...	...	...	...	...	...
39657	Fenway	35307358	Perfect location! Fenway, Newbury st, MFA & More	61151010	Ben	None	42.34525
39658	Dorchester	35386839	Cozy and Spacious #1	266400490	Gil	None	42.30034
39659	Dorchester	35424411	Cozy and Spacious #3	266400490	Gil	None	42.30135
39660	Allston	35513504	Allston & Brighton, Cozy Room Next to Green Li...	266658174	Derya	None	42.34935
39661	Allston	35547817	Allston village, Boston University, near Harvard	30489899	Home On Blaine	None	42.36031

39662 rows × 23 columns

In [7]:
```
reviews_url = "data/Boston/2018/July/reviews.csv"
listings_url = "data/Boston/2018/July/listings.csv"
boston_neighbourhood_2018 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2018.toPandas()
```

Out[7]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Allston	10758	Budget Single - share bath	38440	The Farrington Inn	None	42.
1	Allston	10758	Budget Single - share bath	38440	The Farrington Inn	None	42.
2	Allston	10807	Budget Double - share bath	38440	The Farrington Inn	None	42.
3	Allston	10809	Private Room with 2 Beds share Bath	38440	The Farrington Inn	None	42.
4	Allston	10809	Private Room with 2 Beds share Bath	38440	The Farrington Inn	None	42.
...	...	...	...	...	...	...	
31407	Fenway	26858288	Boston Multi-unit building	34768507	Maryam	None	42.
31408	Fenway	26858288	Boston Multi-unit building	34768507	Maryam	None	42.
31409	Allston	26874729	Temporary Accommodation For August Available.	202133335	Goutham	None	42.
31410	Allston	26874729	Temporary Accommodation For August Available.	202133335	Goutham	None	42.

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
31411	Allston	26892773	Entire Cozy APT, the Best Location	45942122	潇	None	4

31412 rows × 23 columns

In [8]:
```
reviews_url = "data/Boston/2017/October/reviews.csv"
listings_url = "data/Boston/2017/October/listings.csv"
boston_neighbourhood_2017 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2017.toPandas()
```

Out[8]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group
0	Allston	12233830	Modern Boston 2+ Bdrm Harvard/BU/BC/Boston Lan...	6933825	Estee	None
1	Allston	12233830	Modern Boston 2+ Bdrm Harvard/BU/BC/Boston Lan...	6933825	Estee	None
2	Allston	12233830	Modern Boston 2+ Bdrm Harvard/BU/BC/Boston Lan...	6933825	Estee	None
3	Allston	12233830	Modern Boston 2+ Bdrm Harvard/BU/BC/Boston Lan...	6933825	Estee	None
4	Allston	12233830	Modern Boston 2+ Bdrm Harvard/BU/BC/Boston Lan...	6933825	Estee	None
...	...	...	...	...	...	...
20692	Dorchester	533114	The Sawyer | 2BR 2.5BA | 1st Floor	71783	Lance	None
20693	Dorchester	533114	The Sawyer | 2BR 2.5BA | 1st Floor	71783	Lance	None
20694	Dorchester	533114	The Sawyer | 2BR 2.5BA | 1st Floor	71783	Lance	None
20695	Dorchester	533114	The Sawyer | 2BR 2.5BA | 1st Floor	71783	Lance	None
20696	Dorchester	533114	The Sawyer | 2BR 2.5BA | 1st Floor	71783	Lance	None

20697 rows × 23 columns

In [9]:
```
reviews_url = "data/Boston/2016/September/reviews.csv"
listings_url = "data/Boston/2016/September/listings.csv"
boston_neighbourhood_2016 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2016.toPandas()
```

Out[9]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Dorchester	7246272	Stylish Bedroom Suite with Bathroom - Near Subway	19250407	Alex (And Bob)	None	42.317756(
1	Dorchester	7246272	Stylish Bedroom Suite with Bathroom - Near Subway	19250407	Alex (And Bob)	None	42.317756(
2	Dorchester	7246272	Stylish Bedroom Suite with Bathroom - Near Subway	19250407	Alex (And Bob)	None	42.317756(
3	Dorchester	7246272	Stylish Bedroom Suite with Bathroom - Near Subway	19250407	Alex (And Bob)	None	42.317756(
4	Dorchester	7246272	Stylish Bedroom Suite with Bathroom - Near Subway	19250407	Alex (And Bob)	None	42.317756(
...	...	...	...	...	...	...	
13178	Allston	7462268	Allston, close to Harvard Business School + BU D	19548932	Joe	None	42.3581978
13179	Allston	7462268	Allston, close to Harvard Business School + BU D	19548932	Joe	None	42.3581978
13180	Allston	7462268	Allston, close to Harvard Business School + BU D	19548932	Joe	None	42.3581978

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
13181	Allston	7462268	Allston, close to Harvard Business School + BU D	19548932	Joe	None	42.3581978
13182	Allston	7462268	Allston, close to Harvard Business School + BU D	19548932	Joe	None	42.3581978

13183 rows × 23 columns

In [10]:
```
reviews_url = "data/Boston/2015/October/reviews.csv"
listings_url = "data/Boston/2015/October/listings.csv"
boston_neighbourhood_2015 = get_complete_boston_data(listings_url, reviews_
boston_neighbourhood_2015.toPandas()
```

Out[10]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.351929
1	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.351929
2	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.351929
3	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.351929
4	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.351929
...	...	...	...	...	...	...	
8365	Dorchester	3394437	Beautiful Single-Family Boston Home	16912440	Meaghan	None	42.3139552
8366	Dorchester	3394437	Beautiful Single-Family Boston Home	16912440	Meaghan	None	42.3139552
8367	Dorchester	533152	The Penthouse | 2BR 1BA | 4th Floor	71783	Lance	None	42.3218894
8368	Dorchester	533152	The Penthouse | 2BR 1BA | 4th Floor	71783	Lance	None	42.3218894
8369	Dorchester	533152	The Penthouse | 2BR 1BA | 4th Floor	71783	Lance	None	42.3218894

8370 rows × 23 columns

Now that we have the combined data for all 5 years for the required neighbourhoods, we can go into a deepdive on each of the neighbourhoods

Fenway

Let's combine Fenway data from all the years

In [11]:
```
fenway_neighbourhood = boston_neighbourhood_2015.where("neighbourhood == 'F
fenway_neighbourhood = fenway_neighbourhood.union(boston_neighbourhood_2017
fenway_neighbourhood = fenway_neighbourhood.union(boston_neighbourhood_2018
fenway_neighbourhood = fenway_neighbourhood.union(boston_neighbourhood_2019
fenway_neighbourhood = fenway_neighbourhood.union(boston_neighbourhood_2020
fenway_neighbourhood.toPandas()
```

Out[11]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
1	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
2	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
3	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
4	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
...	...	...	...	...	...	...	
24687	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24688	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24689	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24690	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	

	neighbourhood	id	name	host_id	host_name	neighbourhood_group
24691	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None

24692 rows × 23 columns

Get the setiment score for each review using textblob library, but before that, the reviews need to be translated to english, as not all the reviews are in english

In [12]:
```
def translate(sentence):
    try:
        str(TextBlob(x).translate(to= 'en'))
    except:
        return sentence
```

In [13]:
```
translate_udf = F.udf(lambda sentence: translate(sentence) if sentence is n
```

In [14]:
```
fenway_neighbourhood = fenway_neighbourhood.withColumn('comments', translat
fenway_neighbourhood.toPandas()
```

Out[14]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
1	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
2	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
3	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
4	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
...	...	...	...	...	...	...	
24687	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24688	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24689	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24690	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24691	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	

24692 rows × 23 columns

Sentiment Score:

In [11]:
```
sentiment_udf = F.udf(lambda x:TextBlob(x).sentiment.polarity if x is not N
```

In [16]:
```
fenway_neighbourhood = fenway_neighbourhood.withColumn('review_polarity', s
fenway_neighbourhood.toPandas()
```

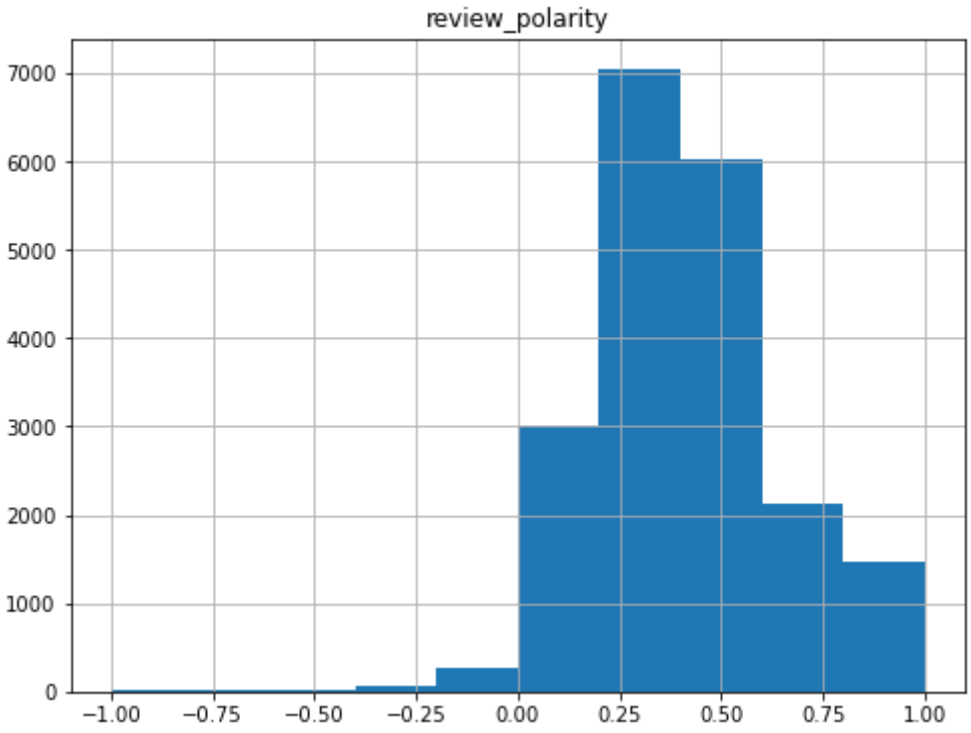
Out[16]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
1	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
2	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
3	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
4	Fenway	5825829	Brownstone Luxury Penthouse Condo	30239904	Andrew	None	42.3466
...	...	...	...	...	...	...	
24687	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24688	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24689	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24690	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	
24691	Fenway	41215360	Private room with fireplace in Center of Boston	322016118	Jay	None	

24692 rows × 24 columns

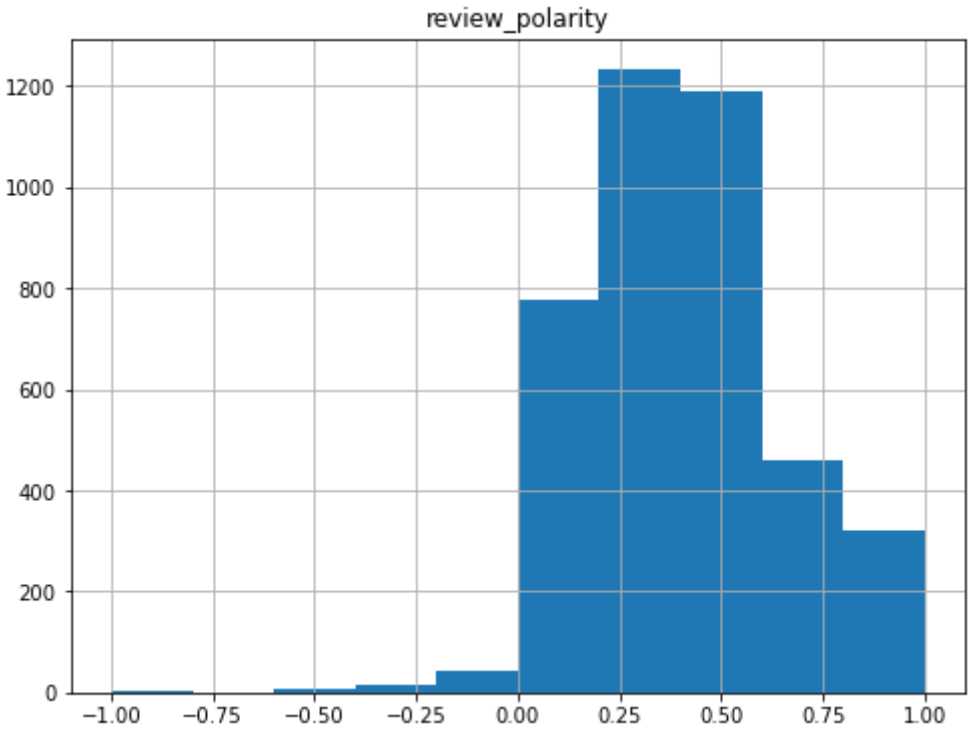
In [17]:
```
fenway_neighbourhood_home = fenway_neighbourhood.where("room_type == 'Entir
fenway_neighbourhood_home = fenway_neighbourhood_home.withColumn("review_po
fenway_neighbourhood_home.hist(column="review_polarity",figsize=(8,6))
```

Out[17]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f748bc86b50
>]],
      dtype=object)
```



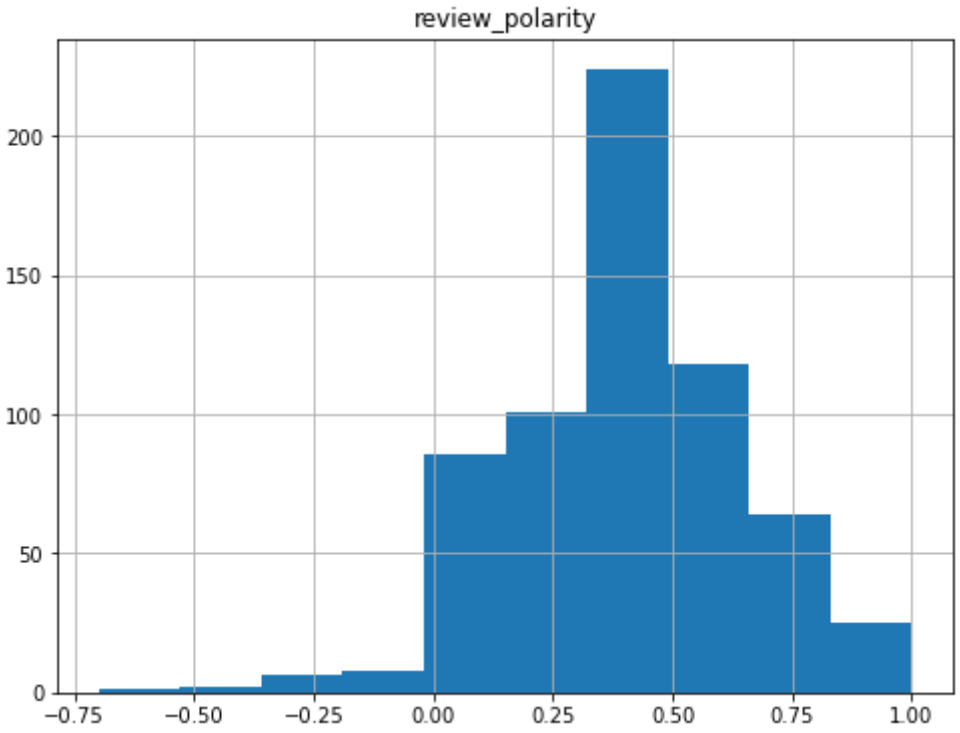
In [18]:
```
fenway_neighbourhood_private = fenway_neighbourhood.where("room_type == 'Pr
fenway_neighbourhood_private = fenway_neighbourhood_private.withColumn("rev
fenway_neighbourhood_private.hist(column="review_polarity", figsize=(8,6))
```

Out[18]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f748a143790
>]],
      dtype=object)
```



In [19]:
```
fenway_neighbourhood_shared = fenway_neighbourhood.where("room_type == 'Sha
fenway_neighbourhood_shared = fenway_neighbourhood_shared.withColumn("revie
fenway_neighbourhood_shared.hist(column="review_polarity", figsize=(8,6))
```

Out[19]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f748a2bbf10
>]],
      dtype=object)
```



Allston:

Let's combine Allston data from all the years

In [20]:
```
allston_neighbourhood = boston_neighbourhood_2015.where("neighbourhood == '
allston_neighbourhood = allston_neighbourhood.union(boston_neighbourhood_20
allston_neighbourhood = allston_neighbourhood.union(boston_neighbourhood_20
allston_neighbourhood = allston_neighbourhood.union(boston_neighbourhood_20
allston_neighbourhood = allston_neighbourhood.union(boston_neighbourhood_20
allston_neighbourhood.toPandas()
```

Out[20]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.35
1	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.35
2	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.35
3	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.35
4	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.35
...	...	...	...	...	...	...	
32266	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32267	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32268	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32269	Allston	41949153	Walk to Harvard ★FREE PARKING★ Clean & Conven...	8441226	Justin	None	

	neighbourhood	id	name	host_id	host_name	neighbourhood_group
32270	Allston	41949153	Walk to Harvard ★FREE PARKING ★ Clean & Conven...	8441226	Justin	None

32271 rows × 23 columns

Let's translate all the reviews to english first

In [21]:
```
allston_neighbourhood = allston_neighbourhood.withColumn('comments', transl
allston_neighbourhood.toPandas()
```

Out[21]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
1	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
2	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
3	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
4	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
...	...	...	...	...	...	...	
32266	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32267	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32268	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32269	Allston	41949153	Walk to Harvard ★FREE PARKING ★ Clean & Conven...	8441226	Justin	None	
32270	Allston	41949153	Walk to Harvard ★FREE PARKING ★ Clean & Conven...	8441226	Justin	None	

32271 rows × 23 columns

Calculate the polarity of each review

In [22]:
```
allston_neighbourhood = allston_neighbourhood.withColumn('review_polarity',
allston_neighbourhood.toPandas()
```

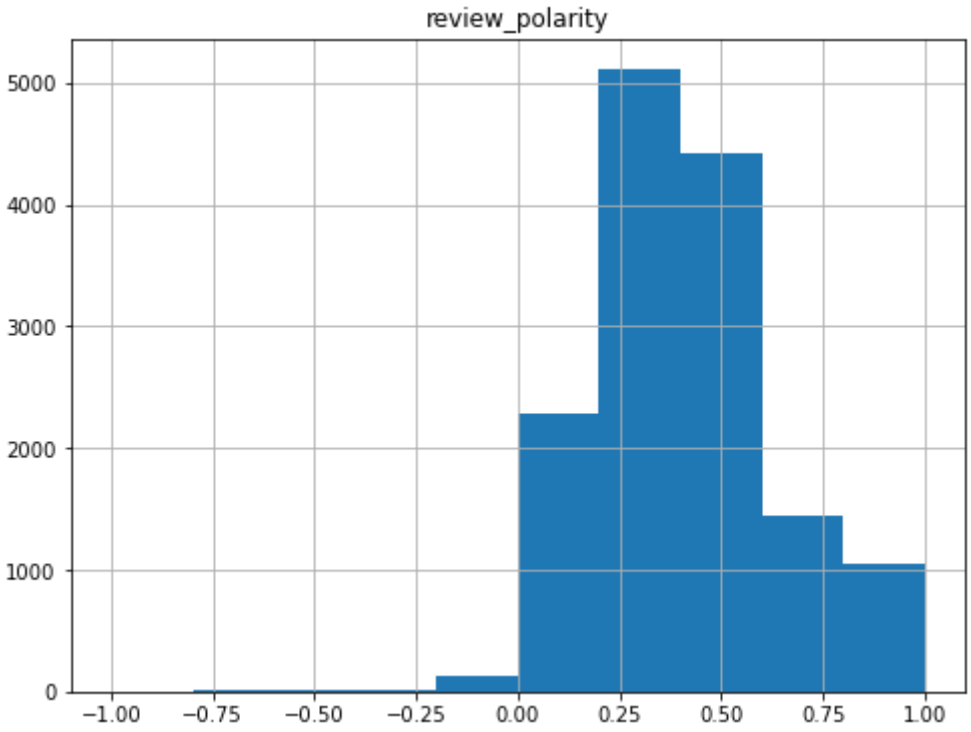
Out[22]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
1	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
2	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
3	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
4	Allston	4278873	STUNNING ROOM PRIVATE BATH ALLSTON	9189449	Cindy	None	42.3
...	...	...	...	...	...	...	
32266	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32267	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32268	Allston	41214604	Cozy, private room in the heart of Boston/Allston	322016118	Jay	None	
32269	Allston	41949153	Walk to Harvard ★FREE PARKING ★ Clean & Conven...	8441226	Justin	None	
32270	Allston	41949153	Walk to Harvard ★FREE PARKING ★ Clean & Conven...	8441226	Justin	None	

32271 rows × 24 columns

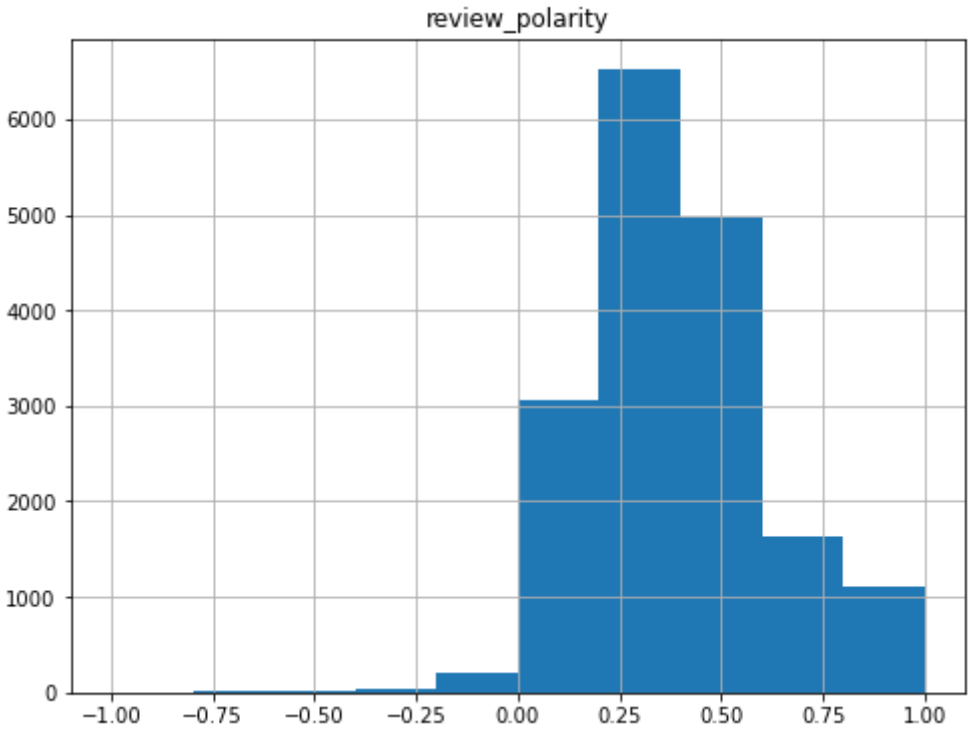
In [23]:
```
allston_neighbourhood_home = allston_neighbourhood.where("room_type == 'Ent
allston_neighbourhood_home = allston_neighbourhood_home.withColumn("review_
allston_neighbourhood_home.hist(column="review_polarity", figsize=(8,6))
```

Out[23]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f7486033190
>]],
      dtype=object)
```



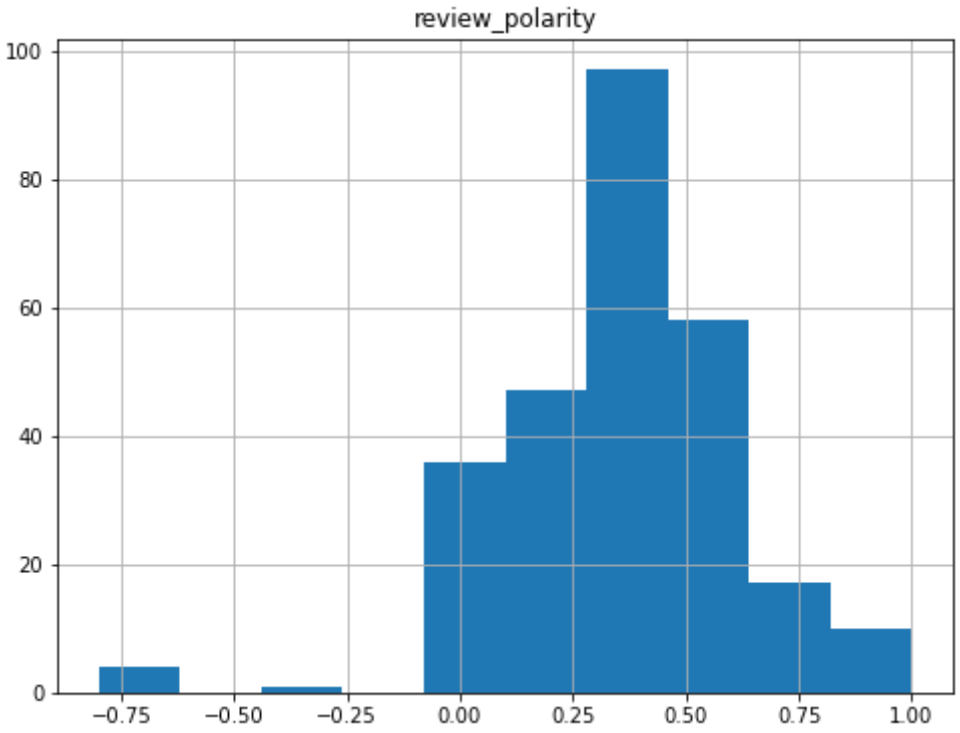
In [24]:
```
allston_neighbourhood_private = allston_neighbourhood.where("room_type == '
allston_neighbourhood_private = allston_neighbourhood_private.withColumn("r
allston_neighbourhood_private.hist(column="review_polarity", figsize=(8,6))
```

Out[24]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f747e691950
>]],
      dtype=object)
```



In [25]:
```
allston_neighbourhood_shared = allston_neighbourhood.where("room_type == 'S
allston_neighbourhood_shared  = allston_neighbourhood_shared .withColumn("r
allston_neighbourhood_shared.hist(column="review_polarity", figsize=(8,6))
```

Out[25]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f747fcede10
>]],
      dtype=object)
```



Dorchester:

Let's combine Dorchester data from all the years

In [12]:
```
dorchester_neighbourhood = boston_neighbourhood_2015.where("neighbourhood =
dorchester_neighbourhood = dorchester_neighbourhood.union(boston_neighbourh
dorchester_neighbourhood = dorchester_neighbourhood.union(boston_neighbourh
dorchester_neighbourhood = dorchester_neighbourhood.union(boston_neighbourh
dorchester_neighbourhood = dorchester_neighbourhood.union(boston_neighbourh
dorchester_neighbourhood.toPandas()
```

Out[12]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
1	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
2	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
3	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
4	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
...	...	...	...	...	...	...	
85446	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	
85447	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	
85448	Dorchester	42831086	Lovely Home | near T	161724560	David	None	
85449	Dorchester	42831086	Lovely Home | near T	161724560	David	None	

	neighbourhood	id	name	host_id	host_name	neighbourhood_group
85450	Dorchester	42831086	Lovely Home | near T	161724560	David	None

85451 rows × 23 columns

Calculate the polarity of each review

In [13]:
```
dorchester_neighbourhood = dorchester_neighbourhood.withColumn('review_pola
dorchester_neighbourhood.toPandas()
```

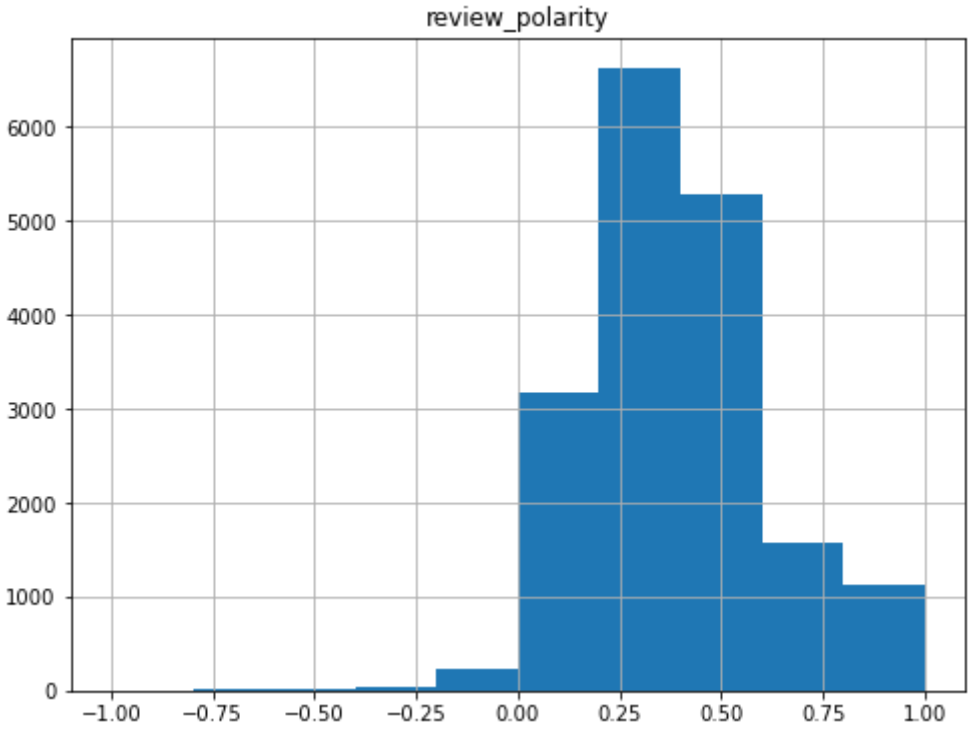
Out[13]:

	neighbourhood	id	name	host_id	host_name	neighbourhood_group	
0	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
1	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
2	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
3	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
4	Dorchester	1178371	Historical Gaslight Street - R2	6430732	Victoria	None	42.294567
...	...	...	...	...	...	...	
85446	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	
85447	Dorchester	42830725	Upscale Duplex | Near T	161724560	David	None	
85448	Dorchester	42831086	Lovely Home | near T	161724560	David	None	
85449	Dorchester	42831086	Lovely Home | near T	161724560	David	None	
85450	Dorchester	42831086	Lovely Home | near T	161724560	David	None	

85451 rows × 24 columns

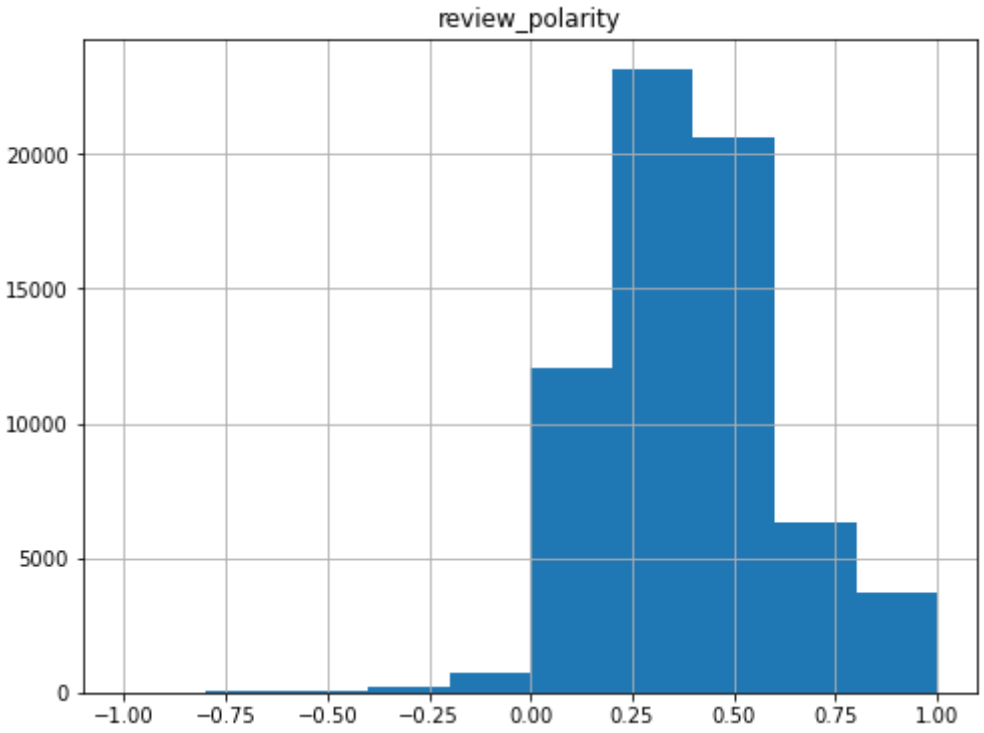
In [17]:
```
dorchester_neighbourhood_home = dorchester_neighbourhood.where("room_type =
dorchester_neighbourhood_home  = dorchester_neighbourhood_home.withColumn("
dorchester_neighbourhood_home.hist(column="review_polarity", figsize=(8,6))
```

Out[17]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f39b4d50090
>]],
      dtype=object)
```



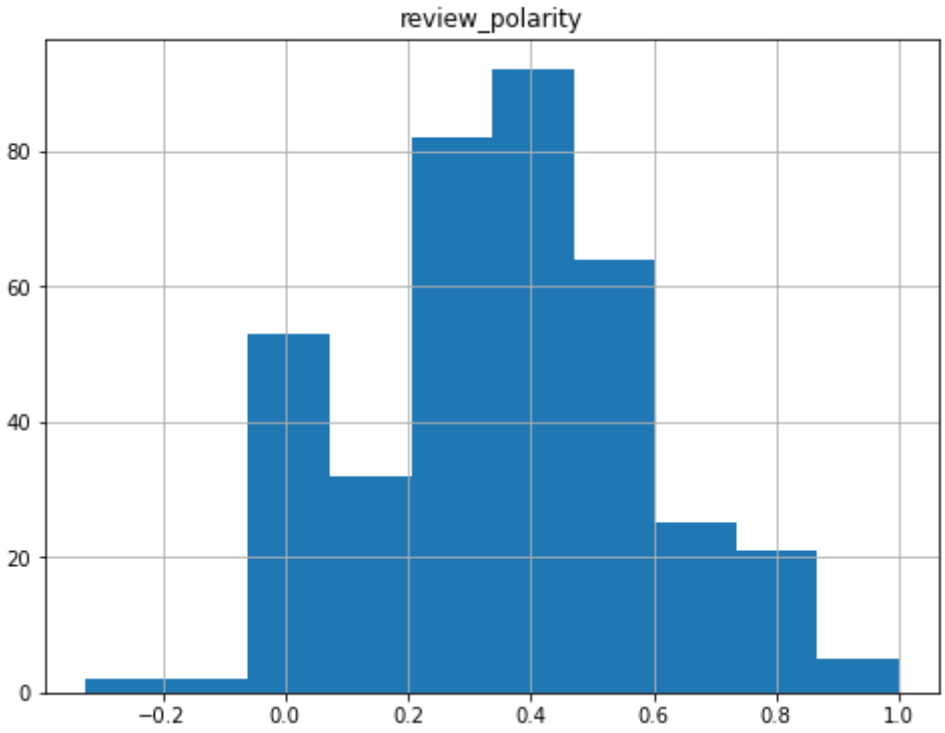
In [18]:
```
dorchester_neighbourhood_private = dorchester_neighbourhood.where("room_typ
dorchester_neighbourhood_private  = dorchester_neighbourhood_private.withCo
dorchester_neighbourhood_private.hist(column="review_polarity", figsize=(8,
```

Out[18]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f399eae6450
>]],
      dtype=object)
```



In [14]:
```
dorchester_neighbourhood_shared = dorchester_neighbourhood.where("room_type
dorchester_neighbourhood_shared  = dorchester_neighbourhood_shared.withColu
dorchester_neighbourhood_shared.hist(column="review_polarity", figsize=(8,6
```

Out[14]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f6dfa2b2c50
>]],
      dtype=object)
```



In [ ]: