Brian Wei
Aaron Lee
BIS 324 Final

Business Problem 1 –

How Have Sales Numbers Fluctuated over the Years?

Northampton County is interested in its real estate market and would like to know how it has changed over the years. Northampton County would like to know the volume of sales and total sales dollars per year. Additionally, Northampton County would like to know how many individual owners participate in the real estate market each year. The county also requested that the most common school district for the purchases be reported as well. Finally, the County requested the average base taxation on the houses sold each year. This data would allow Northampton County to better understand the nature of their real estate market and create a baseline to track future years by, as well as knowing which school districts were popular for new homeowners each year.

SQL

```
select
      distinct(to_char(OH1.sale_date, 'yyyy')) "Year",
      count(OH1.parcel_parcel_id) "Total Number of Sales",
      To_char(sum(OH1.Sale_price), '$999,999,999') "Total Sales Value",
      count(distinct(OH1.owner_owner_id)) "Number of Market Participants",
      stats_mode(P1.School_dist) "Most Popular School District",
      To_char(Round(Avg(T1.base),2), '$999.99') "Average Taxation on Sold Properties"
from owner_history OH1
Join parcel P1 on P1.parcel_id=OH1.parcel_parcel_id
Join Taxes T1 on T1.parcel_parcel_id=OH1.parcel_parcel_id
group by to_char(OH1.sale_date, 'yyyy')
order by to_char(OH1.sale_date, 'yyyy') desc
```

Output

Year	Total Number of Sales	Total Sales Value	Number of Market Participants	Most Popular School District	Average Taxation on Sold Properties
2021	1	$503,201.00	1	Bethlehem School Dist	$774.10
2020	1	$801,983.00	1	Bethlehem School DIST	$475.00
2019	4	$3,933,423.00	3	Bethlehem School DIST	$656.25
2018	5	$2,109,057.00	4	Northhampton School DIST	$496.87
2017	3	$1,118,942.00	3	Bethlehem School Dist	$510.77
2016	3	$1,076,531.00	2	Bethlehem School DIST	$794.92
2015	1	$1,024,833.00	1	Bethlehem School Dist	$526.00
2014	3	$1,656,261.00	2	Bethlehem School DIST	$522.02
2013	3	$1,317,890.00	1	Northhampton School DIST	$566.67
2011	2	$2,047,166.00	1	Northhampton School DIST	$575.00
2010	1	$650,000.00	1	Bethlehem School Dist	$389.99
2009	1	$392,038.00	1	Bangor School DIST	$333.33
2008	4	$1,980,382.00	4	Bangor School DIST	$572.61
2007	3	$2,050,000.00	2	Northhampton School DIST	$641.37
2006	2	$1,098,932.00	2	Bethlehem School Dist	$526.00
2004	1	$123,456.00	1	Bethlehem School Dist	$650.00
2001	1	$768,000.00	1	Bethlehem School Dist	$455.39
1999	1	$512,093.00	1	Bethlehem School Dist	$356.30
1997	2	$937,034.00	2	Bethlehem School Dist	$627.05
1990	1	$1,000,460.00	1	Pen Argyl School DIST	$734.76
1989	1	$425,934.00	1	Bethlehem School Dist	$389.99

Analysis

Within this dataset, the highest volume 2 year period was 2018 and 2019, with 9 combined property sales and over seven million dollars in sales values. The low performance in 2021 and 2020 can be attributed to the COVID-19 pandemic, its variances, and its economic impact. The low sales in 2021 can be further attributed to the lag time between the date of sale and the date that the sale was recorded. It is possible that sales could have happened and could happen before the end of 2021, but will not be recorded until 2022. There are examples of sales being recorded multiple months after the sale has happened.

It should be noted that the SQL query, based on the Owner history table, omits any years where sales were not recorded. Due to the size and limited distribution of our dataset, many years are omitted between 1989 and 2021 (Specifically we are missing 2012, 2005, 2003, 2002, 2000, 1998, and 1991-1996). The gaps in years would not exist if this query was used with a more complete or fully complete dataset.

Business Problem 2 –

Does Northampton County's valuation process need to be improved?

Northampton County wants to know whether their valuation for parcels has been accurate or if the process needs to be improved. The Northampton County parcel database shows the most current valuation for all properties as well as its assessed total value. Using this information, Northampton County would like to determine whether parcels are overvalued, undervalued, or accurate (with a $10,000 margin) in relation to the assessed valuation. Based on the results, the county can verify whether or not their valuation process needs improvement.

SQL

```
SELECT p.parcel_id AS "Parcel ID",
       p.street_number || ' ' ||p.property_location || ', ' || p.city || ' ' || p.state || ', ' || p.zip AS "Address",
       TO_CHAR(v.curr_land + v.curr_bldg, '$999,999') AS "Current Total Value",
       TO_CHAR(v.assessed_land + v.assessed_bldg, '$999,999') AS "Assessed Total Value",
       CASE
           WHEN ((v.assessed_land + v.assessed_bldg) - (v.curr_land + v.curr_bldg)) > 10000
               THEN 'Overvalued'
           WHEN  ((v.curr_land + v.curr_bldg) - (v.assessed_land + v.assessed_bldg)) > 10000
               THEN 'Undervalued'
           ELSE 'Accurate'
       END AS "Assessed Valuation Accuracy"
FROM parcel p JOIN valuation v ON p.parcel_id = v.parcel_parcel_id
ORDER BY p.property_location
```

Output

Parcel ID	Address	Current Total Value	Assessed Total Value	Assessed Valuation Accuracy
P6SW3C10110237	82 4th St, Bethlehem PA, 18015	$95,222	$85,760	Accurate
P6SW3C10110267	450 Birkel Ave, Bethlehem PA, 18015	$105,703	$61,700	Undervalued
P6SW3C10110204	433 Carlton Ave, Bethlehem PA, 18016	$64,770	$38,400	Undervalued
P6SW8A 22 16 0987	626 Carlton Ave, Bethlehem PA, 18015	$94,000	$90,000	Accurate
A4NW4A 2 1 1458	860 Crescent Ln, Allentown PA, 18104	$30,127	$41,573	Overvalued
B7SW8A 2 1 0638	806 East 5th St, Bethlehem PA, 18015	$54,873	$54,800	Accurate
P7SE1C10110287	367 East 7th St, Bethlehem PA, 18015	$48,500	$55,750	Accurate
E9SW4A 2 1 0857	435 Fleckman St, Allentown PA, 18104	$209,000	$219,893	Overvalued
E5HG4A 2 1 0748	728 Hamstead Dr, Allentown PA, 18104	$61,300	$30,700	Undervalued
E8SW5H 2 1 0638	804 Hightower Ave, Bethlehem PA, 18015	$65,334	$52,235	Undervalued
S8SE9C 2 1 0645	450 Lincoln St, Bethlehem PA, 18015	$74,000	$67,153	Accurate
P6SW3C10114692	100 Main St, Bethlehem PA, 18015	$156,000	$156,000	Accurate
P6SE1C10110298	120 Martel St, Bethlehem PA, 18015	$123,000	$133,000	Accurate
D9SE1B 1 9 0128	409 Martino St, Bethlehem PA, 18015	$37,660	$58,000	Overvalued
P6SW3C10826494	50 Packer Ave, Bethlehem PA, 18015	$35,340	$70,953	Overvalued
P6SW3C10110213	30 Polk St, Bethlehem PA, 18015	$57,860	$67,510	Accurate
P6SW3C10110206	515 Selfridge St, Bethlehem PA, 18015	$54,870	$36,809	Undervalued
P6SW3C10110205	123 Smith St, Bethlehem PA, 18015	$80,000	$86,000	Accurate
P6SE1C10110287	12 Washington St, Bethlehem PA, 18015	$91,396	$79,261	Undervalued
E8SW4A 2 1 0638	Woodland Ave, Allentown PA, 18104	$62,800	$71,900	Accurate

Analysis

First, the output above is ordered by Parcel ID. The address column is a concatenation of street number, property location/street name, city, state, and zip code. The 'Current Total Value' and 'Assessed Total Value' are both fairly simple additions of current land and current building values as well as assessed land and assessed building values. The TO_CHAR function was used to apply formatting to the values. Finally, the 'Assessed Valuation Accuracy' column used a CASE statement where 'Undervalued' would be the output if the assessed total value was less than the current total value and 'Overvalued' was the output if the opposite occurred. 'Accurate' was output if the values were within the $10,000 margin established before.

After analyzing the results, it's clear that there is a mix of accurate, overvalued, and undervalued parcels within the county's database. Out of the twenty data points we have, exactly half of the parcels were within the $10,000 margin, meeting the 'accurate' criteria established beforehand. However, six parcels were undervalued and four were overvalued Thus, the conclusion can be made that Northampton County's valuation process for its parcels can definitely be improved upon. From this data, there is no clear indication of whether properties are overvalued more than undervalued or vice versa, but with more parcel data this can become evident.

Business Problem 3 –

Is square footage or the number of bathrooms on a property a better determinant of value?

Square footage and the number of bathrooms are both widely known as important factors when looking at a piece of property. Parcel owners and realtors alike want to know whether square footage or the number of bathrooms is a stronger determinant of value. The current total value of each parcel will be compared to every 1,000 square feet as well as the number of bathrooms (1 per full bath, 0.5 per half bath).

SQL

```
SELECT p.parcel_id AS "Parcel ID",
       TO_CHAR(v.curr_land + v.curr_bldg, '$999,999.99') AS "Current Total Value",
       r.total_sq_ft AS "Total Square Feet",
       TO_CHAR(((v.curr_land + v.curr_bldg) / r.total_sq_ft) * 1000, '$999,999.99') AS "Current Total Value Per 1,000 Sq Ft",
       r.full_baths + (r.half_baths * 0.5) AS "# of Bathrooms",
       TO_CHAR((v.curr_land + v.curr_bldg) / (r.full_baths + (r.half_baths * 0.5)), '$999,999.99') AS "Current Total Value Per Bathroom",
       CASE
           WHEN (((v.curr_land + v.curr_bldg) / r.total_sq_ft) * 1000) > ((v.curr_land + v.curr_bldg) / (r.full_baths + (r.half_baths * 0.5)))
           THEN 'Square Feet'
       ELSE 'Bathrooms'
       END AS "Stronger Determinant of Value"
FROM parcel p JOIN valuation v ON p.parcel_id = v.parcel_parcel_id
          JOIN residential r ON p.parcel_id = r.parcel_parcel_id
WHERE r.full_baths + (r.half_baths * 0.5) > 0
ORDER BY "Current Total Value" DESC
```

Output

Parcel ID	Current Total Value	Total Square Feet	Current Total Value Per 1,000 Sq Ft	# of Bathrooms	Current Total Value Per Bathroom	Stronger Determinant of Value
P6SW3C10114692	$156,000.00	2400	$65,000.00	1.5	$104,000.00	Bathrooms
P6SE1C10110298	$123,000.00	3000	$41,000.00	2	$61,500.00	Bathrooms
P6SW3C10110267	$105,703.00	2250	$46,979.11	1.5	$70,468.67	Bathrooms
P6SW3C10110237	$95,222.00	2450	$38,866.12	2.5	$38,088.80	Square Feet
P6SW8A 22 16 0987	$94,000.00	2345	$40,085.29	2.5	$37,600.00	Square Feet
P6SE1C10110287	$91,396.00	2600	$35,152.31	2.5	$36,558.40	Bathrooms
P6SW3C10110205	$80,000.00	2000	$40,000.00	2	$40,000.00	Bathrooms
S8SE9C 2 1 0645	$74,000.00	2500	$29,600.00	1.5	$49,333.33	Bathrooms
E8SW5H 2 1 0638	$65,334.00	2400	$27,222.50	2.5	$26,133.60	Square Feet
P6SW3C10110204	$64,770.00	1900	$34,089.47	2.5	$25,908.00	Square Feet
E8SW4A 2 1 0638	$62,800.00	2200	$28,545.45	2.5	$25,120.00	Square Feet
E5HG4A 2 1 0748	$61,300.00	2500	$24,520.00	2.5	$24,520.00	Bathrooms
P6SW3C10110213	$57,860.00	2000	$28,930.00	1	$57,860.00	Bathrooms
B7SW8A 2 1 0638	$54,873.00	3900	$14,070.00	2	$27,436.50	Bathrooms
P6SW3C10110206	$54,870.00	2250	$24,386.67	2.5	$21,948.00	Square Feet
P7SE1C10110287	$48,500.00	3250	$14,923.08	1.5	$32,333.33	Bathrooms
D9SE1B 1 9 0128	$37,660.00	2356	$15,984.72	2.5	$15,064.00	Square Feet
P6SW3C10826494	$35,340.00	2360	$14,974.58	2.5	$14,136.00	Square Feet
A4NW4A 2 1 1458	$30,127.00	2170	$13,883.41	2.5	$12,050.80	Square Feet

Analysis

This output shows Parcel ID, the current total value which is a straightforward addition of current land and current building, total square feet, and the number of bathrooms. The 'Current Total Value Per Bathroom' is found by dividing the current total value of each parcel by every 1,000 square feet. 'Current Total Value Per Bathroom' is similarly calculated by dividing the current total value by the number of bathrooms. The 'Stronger Determinant of Value' column

output is whichever determinant is greater, square feet or bathrooms. In addition, parcels with no bathrooms were excluded from this analysis.

Looking at the 'Stronger Determinant of Value' column, there is no apparent determinant that occurs more often than the other. The number of bathrooms in a property was valued more in ten of the parcels while the value per 1,000 square feet was valued more in the remaining nine parcels. Based on this data and analysis, homeowners and realtors can conclude that to increase the value of their properties, they can focus equally on square footage and the number of bathrooms.

Business Problem 4 –

What are the Rates of Usage and Acreage for Land Use Codes and Act Flags?

Northampton County wants to assess the efficacy of its current land use codes and act flags. Northampton County wants to know the number of parcels that use each act flag and land use code in the county. Additionally, the county would like to know the acreage under each flag or code. The county wants to identify codes and act flags that have no usage or low usage. (Note: The limit for low usage is 3, but if a larger dataset was used, the limit may need to be raised).

SQL

```
select
      'Code '||C1.code "Code or Act Flag",
      C1.Code_desc "Description",
      count (HC1.Residential_parcel_id) "Number of Parcels with Code or Flag",
      sum(P1.CAMA_acres) "Acreage Under Code or Flag",
      case
          when count(HC1.Residential_parcel_id) < 1 then 'No Usage'
          when count(HC1.Residential_parcel_id) < 3 then 'Low Usage'
          else 'Acceptable Usage'
          end "Usage Rate"
from codes C1
join has_code HC1 on HC1.codes_code=C1.code
Join parcel P1 on HC1.Residential_parcel_id=P1.Parcel_id
group by C1.code, C1.Code_desc
union all
select
      'Act Flag '|| A1.act_flag "Code or Act Flag",
      A1.act_desc "Description",
      count(HA1.Parcel_parcel_id) "Number of Parcels with Code or Flag",
      sum(P1.CAMA_acres) "Acreage Under Code or Flag",
      case
          when count(HA1.Parcel_parcel_id) < 1 then 'No Usage'
          when count(HA1.Parcel_parcel_id) < 3 then 'Low Usage'
          else 'Acceptable Usage'
          end "Usage Rate"
from act_flags A1
join has_act HA1 on HA1.act_flags_act_flag=A1.act_flag
Join parcel P1 on HA1.Parcel_parcel_id=P1.Parcel_id
group by A1.act_flag, A1.act_desc
```

Output

Code or Act Flag	Description	Number of Parcels with Code or Flag	Acreage Under Code or Flag	Usage Rate
Code 191	Veterans	5	1.0316	Acceptable Usage
Code 100	Exempt Institutional Vacant	3	2.1268	Acceptable Usage
Code 400	General	4	1.074	Acceptable Usage
Code 180	Improvements	4	1.2073	Acceptable Usage
Code 350	Fraternal Organizations	3	1.4738	Acceptable Usage
Code 102	Landfill	1	1.513	Low Usage
Act Flag LERTA	Economic Revitalization Act	5	3.1876	Acceptable Usage
Act Flag 319	Clean and Green Act	3	.195	Acceptable Usage
Act Flag 43	Agricultural Security Area	3	1.2802	Acceptable Usage
Act Flag 149/4	Agricultural Easement Act	2	.7386	Low Usage
Act Flag KOZ	Keystone Opportunity Zone	5	2.9091	Acceptable Usage
Act Flag 66	Public Utility Realty Tax Act	2	.116	Low Usage

Analysis

      Firstly, it should be noted that the limit we used to determine whether a code or act flag was low usage (3), is possibly much lower than would actually be used by the county. We selected this low number because of the small size of our dataset. The output shows that there are multiple flags and codes that have low usage, but all flags and codes are being used by at least one parcel, as there are no rows returning 'No Usage'. Additionally, the results also show that 'low usage' flags and codes need to be assessed before being discarded. Some codes such as 'Landfill' are probably very important to keep designated because there are likely not very many landfills, so low usage would be expected. Contrasting this, Act Flag 149/4 'Agricultural Easement Act', despite having a higher number of parcels than the landfill code, is a possible candidate for removal, if it was designed or expected to be widely used. It is possible that the Agricultural Easement Act is an older Act Flag that has fallen out of favor, or no longer fits as the county plans to urbanize. This query would highlight any unused or hardly used flags and codes, making it easier for county officials to decide whether to eliminate any codes or flags.

Business Problem 5 –

Averaging Property Values and Data by School District

Realtors and contractors are wondering about the different prices per square foot by school district in Northampton county. Additionally, these companies want to know how average square footage, valuation, and sales price per square foot change by school district. In addition, the realtors and contractors would like to know the average number of bedrooms by school district, to get a better understanding of how many people on average live in a house in each school district. Knowing this information would give them more information as to the number of bedrooms to build in each school district as well as the relative price per square foot between all the school districts.

SQL

```
select
    distinct(P1.school_dist) "School District",
    Round(avg(R1.total_sq_ft),2) "Average Square Feet",
    Round(avg(R1.bedrooms),2) "Average Bedrooms",
    To_char(round(avg(V1.curr_bldg/R1.total_sq_ft),2), '$999.99') "Average Current Price Per Square Foot",
    To_char(round(avg(V1.assessed_bldg/R1.total_sq_ft),2), '$999.99') "Average Assessed Price Per Square Foot",
    To_char(round(avg(OH1.sale_price/R1.total_sq_ft),2), '$999.99') "Average Sale Price Per Square Foot",
    stats_mode(R1.grade) "Most Common Grade",
    count(P1.parcel_id) "Number of Properties"
from parcel P1
join residential R1 on R1.parcel_parcel_id=P1.parcel_id
join valuation V1 on V1.parcel_parcel_id=P1.parcel_id
join Owner_History OH1 on OH1.parcel_parcel_id=P1.parcel_id
group by P1.school_dist
```

Output

School District	Average Square Feet	Average Bedrooms	Average Current Price Per Square Foot	Average Assessed Price Per Square Foot	Average Sale Price Per Square Foot	Most Common Grade	Number of Properties
Bangor School DIST	2356	4	$10.47	$7.64	$244.01	A	3
Bethlehem School DIST	2659.82	4.11	$17.57	$16.91	$225.42	A	28
Northampton School DIST	3750	.5	$21.78	$21.62	$172.30	C	6
Pen Argyl School DIST	2200	5	$20.32	$4.14	$219.60	B	3

Analysis

Looking at the data, we can see that the 4 school districts are fairly similar in square footage and prices and values per square footage, but Northampton, has much fewer bedrooms on average and a much lower common grade.

Ultimately, the data is skewed due to the number of Bethlehem properties and relative lack of other properties. This would be fixed with a larger dataset.