Business Context for Data Analytics

Focus : SQL and Tableau

Final Presentation

Nicole Hodson

1– Business Understanding – High Level

Case Study

- This case study looks at the Trans Mountain Pipeline and the issues surrounding it.

Introduction

- This Presentation will examine the impacts of Pipelines on indigenous communities in Canada and go into Data found about a few pipelines that run through Canada for Exporting Oil.

5W 1H Analysis 1 – Who

#	Question	Answer
Q1.1	Who is involved?	Canadian Federal Government, Tsleil Waututh First Nations, Squamish Nation, Cold water Indian Band, Canadian people, Provincial Governments, Kinder Morgan, The National Energy Board
Q1.2	Who is affected?	Canadian Governments, Indigenous communities, the Environment.
Q1.3	Who will benefit?	Kinder Morgan, The Canadian Government, Energy Companies wanting to use the pipeline to transport their products to the Pacific Rim Market.
Q1.4	Who will be harmed?	The Indigenous Communities, The Environment

5W 1H Analysis 2 – What

#	Question	Answer
Q2.1	What is your topic narrowed down in a simple phrase/sentence?	The Canadian Government pushed through the plans for building a pipeline from Edmonton to the Westrige Marine Terminal in Burnaby, passing through and exiting on Indigenous lands.
Q2.2	What does your topic involve? (i.e. What are the different parts to it?)	The topic involves the disconnect of Canadian Constitutional law and the rights for any Indigenous Community to veto a plan like this due to Stewardship policy and their land rights given to them by the government.
Q2.3	What is it similar to / different from?	Building a pipeline is not new for the Canadian People. Where it differs is the Canadian Government passing laws to change the way that Oil and gas companies can operate under new environmental laws.
Q2.4	What might be affected/changed by your topic?	Location changes would change some of the concerns that the local communities find to be problems.

5W 1H Analysis 3 –When

#	Question	Answer
Q3.1	When does this take place? When did this take place? When will it take place? When should this take place?	This is taking place right now, Aboriginal Communities and other Environmental groups are attempting to fight the Canadian Governments decision to go ahead with the project. The Federal Court of Appeal chose not to listen to them the first time. Now there are many areas along the route that have experienced protests.
Q3.2	Does when this takes place affect the topic?	If this pipeline was to be built 20-40 years ago, I don't think as many people would understand the environmental, social, and economic hardships like we understand them today.

5W 1H Analysis 4 – Where

#	Question	Answer
Q4.1	Where does this take place? (Where did it …. Where will it … Where should it ….?)	The Pipeline runs from Edmonton to an Inlet in Burnaby BC – through a few indigenous lands. It brings in both provinces of Alberta and British Columbia.
Q4.2	Does it matter where it takes place? Is it affected by location?	Location affects some of the impacts, but it does not change the environmental impacts being suggested in the case study. The environmental impacts would still be there, no matter the location of the pipeline. Bitumen is noted to be 30-70% more polluting than traditional oil refining.

5W 1H Analysis 5 – Why

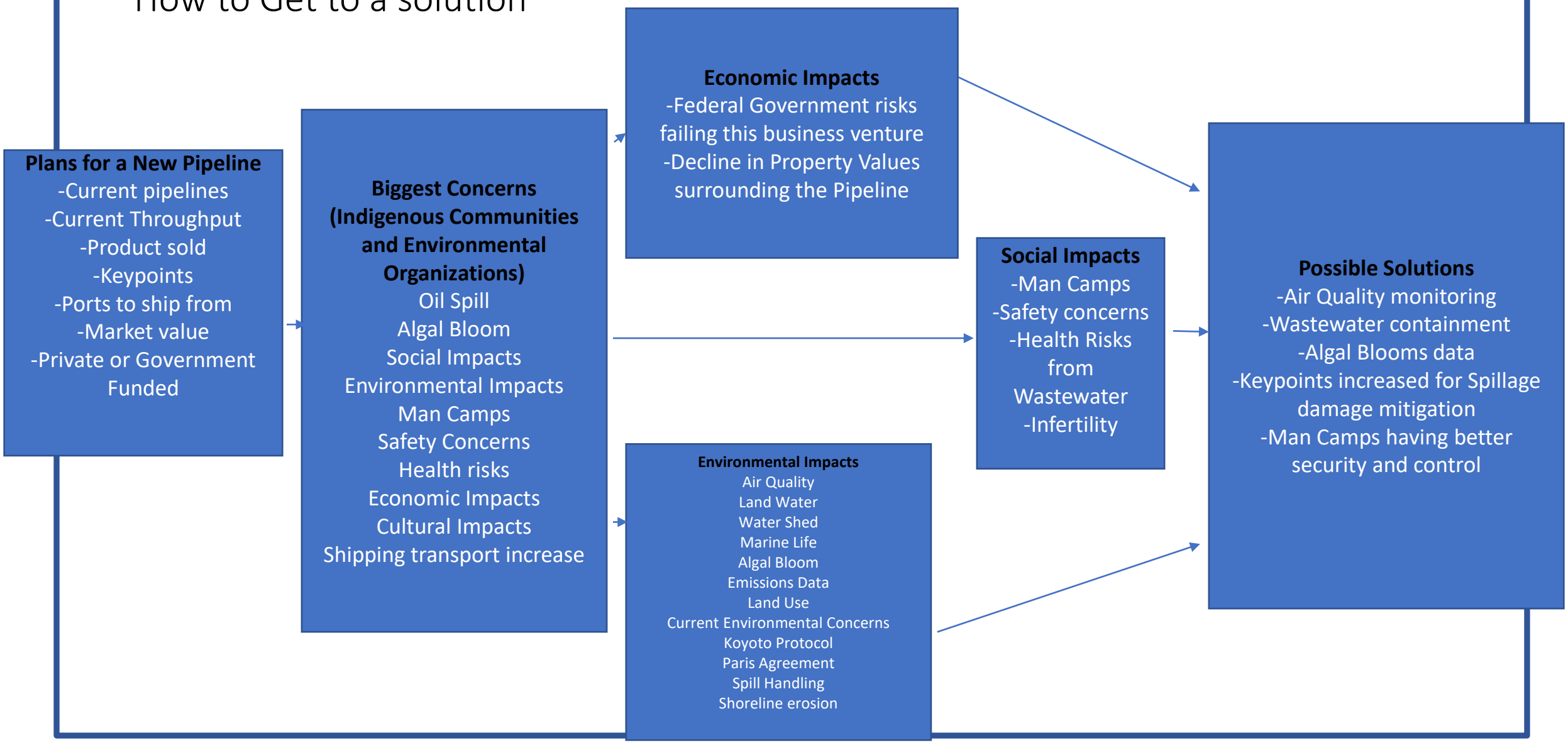
#	Question	Answer
Q5.1	Why is this topic important? Why does it matter?	The Federal Government is attempting to go through with plans, despite the Constitutional rights of indigenous groups with heir Stewardship policies, Aboriginal titles, Free-Prior-Informed Consent (FPIC), United Nations Declaration on the Rights of Indigenous People (UNDRIP), among other rights infringing claims.
Q5.2	Why do certain things happen? (What are some causes and effects within the topic?)	In this case, the Federal Government involked a few amendments to certain consistutional rights, such as one that mentions they can do this if their choices are better for more people than the few that are voicing their concerns.

5W 1H Analysis 6 – How

#	Question	Answer
Q6.1	How does this topic work? How does it function? How does it do what it does?	Canadian Government only see this pipeline as a money opportunity and they have already spent a large amount of money on purchasing the pipeline plans, seeing it as doing the greeater good for the majority of Canadian Citizens.
Q6.2	How did it come to be?	Many Oil companies see pipelines as the only path forward in order to get their product to a larger market, the Pacific Coast Rim.
Q6.3	How are those involved affected?	The Government will lose a lot of money if not inacted. Indigenous communities will experience social, environmental, and cultural issues for their communities. Oil Companies will benefit by having a large market to sell Bitumen to.

2 – Data Understanding – High Level

How to Get to a solution



Data Dictionary

No	Topic	Definition
1	MNP	Canadian Mainline Pipeline
2	Bitumen	A type of crude oil which can be mined
3	NEB	National Energy Board
4	Key Point	Location alone pipeliune where through puts are measured and reported, determined by corporate entity
5	Throughput(1000m3/d)	The volume of product flowing through the pipeline at the Key Point Indicated, in thousand cubic meters per day
6	Trade Type	Export, Import, Intracanada
7	FPIC	Free, prior, informed consent
8	CBD	Convention on Biological Diversity
9	UNDRIP	United Nations Declaration on the Rights of Indigenous People
10	NGO's/Environmental groups	Non-Government Organizations with interests in the health and safety of individuals as well as the environmental impacts of Canadian Oil Businesses

Data Dictionary

No	Topic	Definition
11	CER	Canada's Energy Regulator
12	Section 35 of the Canadian Constitution	Rights to fish, hunt, gather for sustenance, ceremonial and cultural survival of the indigenous peoples of Canada.

1 (Revisited) – Business Understanding – Low Level

Business Questions

- Question 1) What is the amount that each Canadian Pipeline exports in 1000 cubic meters per day for an entire year and where are they located/going?
- Question 2) What is the percentage of throughput compared to capacity of the top 3 exporting pipelines averaged by year?

2 (Revisited) – Data Understanding – Low Level

Data Set(s)

alliance-throughput-and-capacity-dataset

brunswick-throughput-and-capacity-dataset

tcpl-mainline-throughput-and-capacity-dataset

enbridge-mainline-throughput-and-capacity-dataset

foothills-throughput-and-capacity-dataset

keystone-throughput-and-capacity-dataset

mnp-throughput-and-capacity-dataset

tqm-throughput-and-capacity-dataset

trans-mountain-throughput-and-capacity-dataset

westcoast-throughput-and-capacity-dataset

Database Diagram – Tables and Columns

keystone

Column Name	Data Type	Allow Nulls
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(100)	☐
Latitude	float	☐
Longitude	float	☐
Direction	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Product	nvarchar(50)	☐
Throughput_1000_m3_d	numeric(18, 0)	☑
Available_Capacity_1000...	numeric(18, 0)	☑
Reason_For_Variance	nvarchar(200)	☑
		☐

westcoast

Column Name	Data Type	Allow Nulls
Date	tinyint	☐
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_Of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Capacity_1000_m3_d	numeric(18, 0)	☑
Throughput_1000_m3_d	numeric(18, 0)	☑
Throughput_GJ_d	numeric(18, 0)	☑
		☐

Foothills

Column Name	Data Type	Allow Nulls
Date	tinyint	☐
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_Of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Capacity_1000_m3_d	numeric(18, 0)	☐
Throughput_1000_m3_d	numeric(18, 0)	☐
Throughput_GJ_d	numeric(18, 0)	☐
		☐

Enbridge

Column Name	Data Type	Allow Nulls
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Product	nvarchar(50)	☐
Throughput_1000_m3_d	float	☑
Nameplate_Capacity_100...	float	☑
Available_Capacity_1000...	float	☑
		☐

Alliance

Column Name	Data Type	Allow Nulls
Date	tinyint	☐
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Capacity_1000_m3_d	float	☑
Throughput_1000_m3_d	float	☑
Throughput_GJ_d	float	☐
		☐

TCLP

Column Name	Data Type	Allow Nulls
Day	tinyint	☐
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_Of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Capacity_1000_m3_d	float	☑
Throughput_1000_m3_d	float	☑
Throughput_GJ_d	float	☑
		☐

MNP

Column Name	Data Type	Allow Nulls
Date	tinyint	☐
Month	tinyint	☐
Year	smallint	☐
Corporate_Entity	nvarchar(50)	☐
Pipeline_Name	nvarchar(50)	☐
Key_Point	nvarchar(50)	☐
Latitude	float	☐
Longitude	float	☐
Direction_Of_Flow	nvarchar(50)	☐
Trade_Type	nvarchar(50)	☐
Capacity_1000_m3_d	numeric(18, 0)	☑
Throughput_1000_m3_d	numeric(18, 0)	☑
Throughput_GJ_d	numeric(18, 0)	☑
		☐

3 –
Data Preparation

Most data cleansing was done in SQL.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Date	Month	Year	Corpor	Pipelin	Key Po	Latitud	Longitu	Directio	Trade T	Capacit	Throug	Throug	ut (GJ/d)
1	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51717.21	48884.2	1973514	
2	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51733.31	48708.1	1968322	
3	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51748.75	48355.3	1964453	
4	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51763.53	49248.5	1999268	
5	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51777.64	49377.8	2000305	
6	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51791.07	48571.6	1967702	
7	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51803.83	47770.8	1935415	
8	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51815.91	48405.8	1961558	
9	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51827.31	49452	1975602	
10	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51838.02	49155.4	1969785	
11	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51848.04	48324.5	1926070	
12	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51857.36	47855	1914396	
13	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51865.99	51964.7	2086700	
14	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51873.92	50187.3	2015437	
15	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51881.15	50475.4	2029086	
16	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51887.67	49944.8	2006442	
17	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51893.49	49409.6	1985036	
18	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51898.6	49166.2	1975345	
19	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51903.01	45024.4	1808974	
20	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51906.7	50478	2041574	
21	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51909.68	49694.9	2002695	
22	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51911.95	49356.7	1989272	
23	1	2006	Alliance P	Alliance P	Border	49	-101.588	southeast	export	51913.5	49293.5	1987829	

1

Snip of Original data source from Alliance Pipeline dataset. I loaded all tables into SQL and coded to eliminate what I didn't need.

2

This SQL Query allowed me to limit required information, and gather what I needed to answer the questions asked. This shows....

```
SELECT

      [Year]
      ,[Corporate_Entity]
      ,[Pipeline_Name]
      ,[Trade_Type]
      , avg([Throughput_1000_m3_d]/[Capacity_1000_m3_d])*100 'Total Throughput'
  ,[Latitude]
  ,[Longitude]
  ,[Key_Point]

  FROM [dbo].[Alliance]
  where Trade_Type = 'Export' and year >= 2009
  group by Year, Corporate_Entity, Pipeline_Name, Trade_Type, Longitude, Latitude, Key_Point
  order by [Year]
```

Results | Messages

	Year	Corporate_Entity	Pipeline_Name	Trade_Type	Total Throughput	Latitude	Longitude	Key_Point
1	2009	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	93.312456704232	49	-101.587997436523	Border
2	2010	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	94.2940764434248	49	-101.587997436523	Border
3	2011	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.6518965998208	49	-101.587997436523	Border
4	2012	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	90.4569316165674	49	-101.587997436523	Border
5	2013	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.4884565268007	49	-101.587997436523	Border
6	2014	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.0080852803216	49	-101.587997436523	Border
7	2015	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	88.9192878539707	49	-101.587997436523	Border
8	2016	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	92.7542792573846	49	-101.587997436523	Border
9	2017	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	95.4109467259509	49	-101.587997436523	Border
10	2018	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	98.0141819144813	49	-101.587997436523	Border
11	2019	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	98.0061854339463	49	-101.587997436523	Border
12	2020	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	80.0341257639829	49	-101.587997436523	Border

3

I then copied results from the Querys back into Excel to use this specific information from each required data set and combined the information into one excel sheet.

Data Preparation – SQL Query for Question 1

```
select
   [year],[corporate_entity],[pipeline_name],
[trade_type] ,sum([Throughput_1000_m3_d]) 'Total
Throughput'
   from [dbo].[TCLP]
   where Trade_Type = 'Export' and year >= 2006
  group by Year, Corporate_Entity, Pipeline_Name,
Trade_Type
  order by [Year]
```

I changed the table name for every table I needed information from and copied and pasted it together into excel.

Data Preparation – Data set returned for Query 1

Results | Messages

	year	corporate_entity	pipeline_name	trade_type	Total Throughput
1	2006	TransCanada PipeLines Limited	Canadian Mainline	Export	47206065.3509521
2	2007	TransCanada PipeLines Limited	Canadian Mainline	Export	46351047.1903687
3	2008	TransCanada PipeLines Limited	Canadian Mainline	Export	43304275.4860229
4	2009	TransCanada PipeLines Limited	Canadian Mainline	Export	35163505.8632202
5	2010	TransCanada PipeLines Limited	Canadian Mainline	Export	29468107.5021362
6	2011	TransCanada PipeLines Limited	Canadian Mainline	Export	23433337.490388
7	2012	TransCanada PipeLines Limited	Canadian Mainline	Export	15527429.6348349
8	2013	TransCanada PipeLines Limited	Canadian Mainline	Export	12760673.1794277
9	2014	TransCanada PipeLines Limited	Canadian Mainline	Export	14057296.1158107
10	2015	TransCanada PipeLines Limited	Canadian Mainline	Export	11393110.7211742
11	2016	TransCanada PipeLines Limited	Canadian Mainline	Export	13994671.5044007
12	2017	TransCanada PipeLines Limited	Canadian Mainline	Export	17544388.41038
13	2018	TransCanada PipeLines Limited	Canadian Mainline	Export	20155540.768942
14	2019	TransCanada PipeLines Limited	Canadian Mainline	Export	18581113.1211252
15	2020	TransCanada PipeLines Limited	Canadian Mainline	Export	8285616.12591553

Data Preparation – SQL Query for Question 2

```
SELECT

      [Year]
      ,[Corporate_Entity]
      ,[Pipeline_Name]
      ,[Trade_Type]
      , avg([Throughput_1000_m3_d]/[Capacity_1000_m3_d])*100 'Total
Throughput'
  ,[Latitude]
  ,[Longitude]
  ,[Key_Point]

  FROM [dbo].[Alliance]
  where Trade_Type = 'Export' and year >= 2009
  group by Year, Corporate_Entity, Pipeline_Name, Trade_Type,
Longitude, Latitude, Key_Point
  order by [Year]
```

I changed the table name to get different information from each pipeline so I could put them together into and Excel sheet.

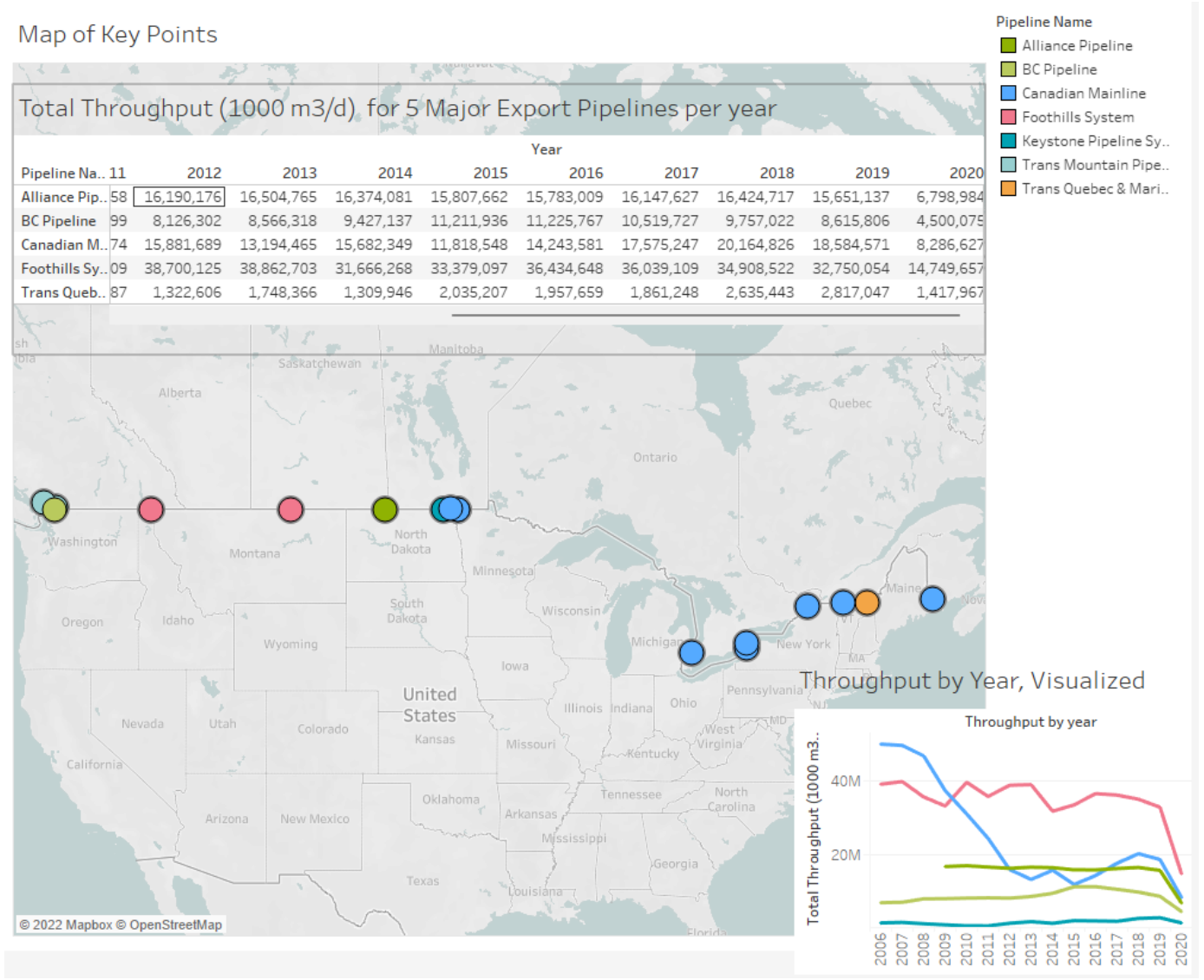
Data Preparation – Data set returned for Query 2

Results | Messages

	Year	Corporate_Entity	Pipeline_Name	Trade_Type	Total Throughput	Latitude	Longitude	Key_Point
1	2009	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	93.312456704232	49	-101.587997436523	Border
2	2010	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	94.2940764434248	49	-101.587997436523	Border
3	2011	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.6518965998208	49	-101.587997436523	Border
4	2012	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	90.4569316165674	49	-101.587997436523	Border
5	2013	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.4884565268007	49	-101.587997436523	Border
6	2014	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	91.0080852803216	49	-101.587997436523	Border
7	2015	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	88.9192878539707	49	-101.587997436523	Border
8	2016	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	92.7542792573846	49	-101.587997436523	Border
9	2017	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	95.4109467259509	49	-101.587997436523	Border
10	2018	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	98.0141819144813	49	-101.587997436523	Border
11	2019	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	98.0061854339463	49	-101.587997436523	Border
12	2020	Alliance Pipeline Limited Partnership	Alliance Pipeline	Export	80.0341257639829	49	-101.587997436523	Border

4 – Modeling

Dashboard for Question 1



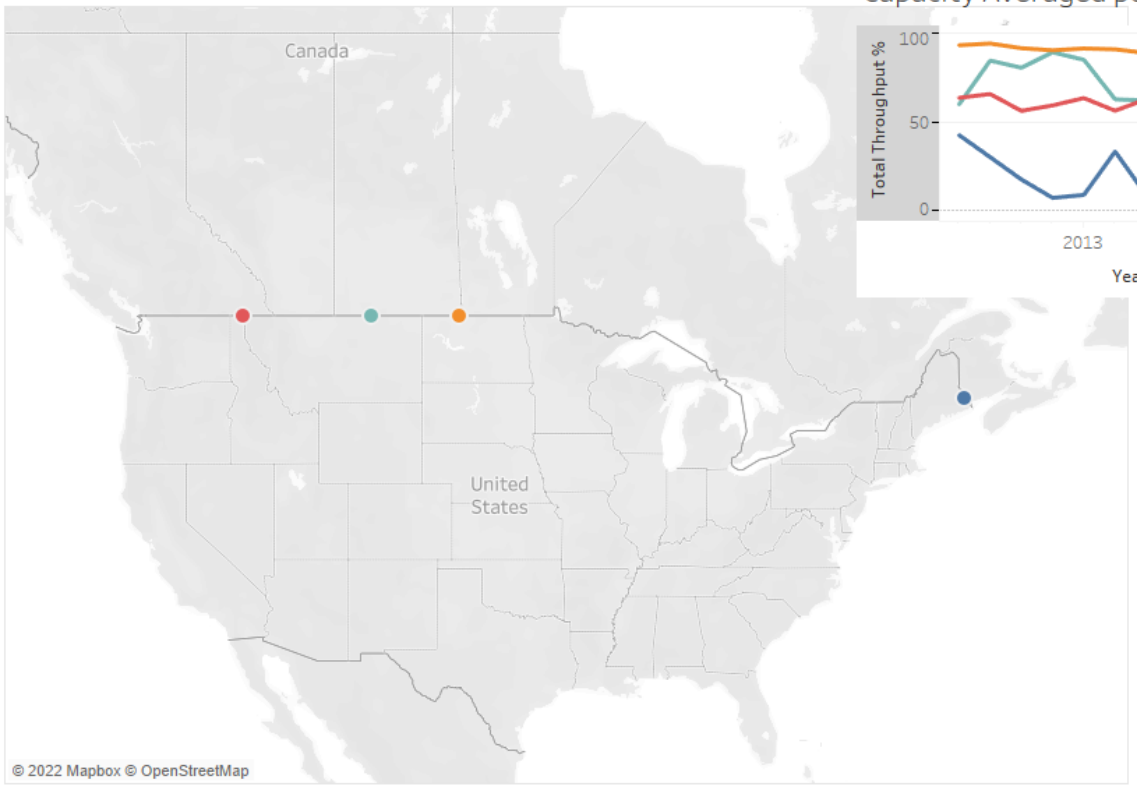
Dashboard For Question 2

Year

Key Point	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Baileyville, ..	42.69	30.21	17.62	7.24	8.89	33.37	8.72	5.08	0.62	0.17	0.05	0.02
Border	93.31	94.29	91.65	90.46	91.49	91.01	88.92	92.75	95.41	98.01	98.01	80.03
Kingsgate	63.64	65.81	56.30	59.39	63.52	56.40	62.88	67.79	70.85	75.86	76.27	75.25
Monchy	59.84	84.66	80.62	89.34	85.11	62.80	62.22	68.66	63.31	51.86	41.94	28.90

Key Point
- Baileyville, Ma. / St. St..
- Border
- Kingsgate
- Monchy

Locations of Top Export Key Points



Throughput Percentage of Capacity Averaged per Year

5 – Evaluation

Conclusion – Answer to Question 1

- From the Data presented, we can understand that exports of oil in Canada have decreased over time. With more data, we can understand why. Factors that could have data to showcase an accurate reasoning could be world market data, world demand for oil, world supply of oil.

Conclusion – Answer to Question 2

- Conclusions can be drawn using this dashboard. We can see that some pipelines functions at a high rate of capacity, while others fall short. More data is needed to understand exactly why. More data suggested could be downtimes of pipelines, spill mitigation, maintenance schedules, construction.