Nobel Prize Data Warehouse

Introduction

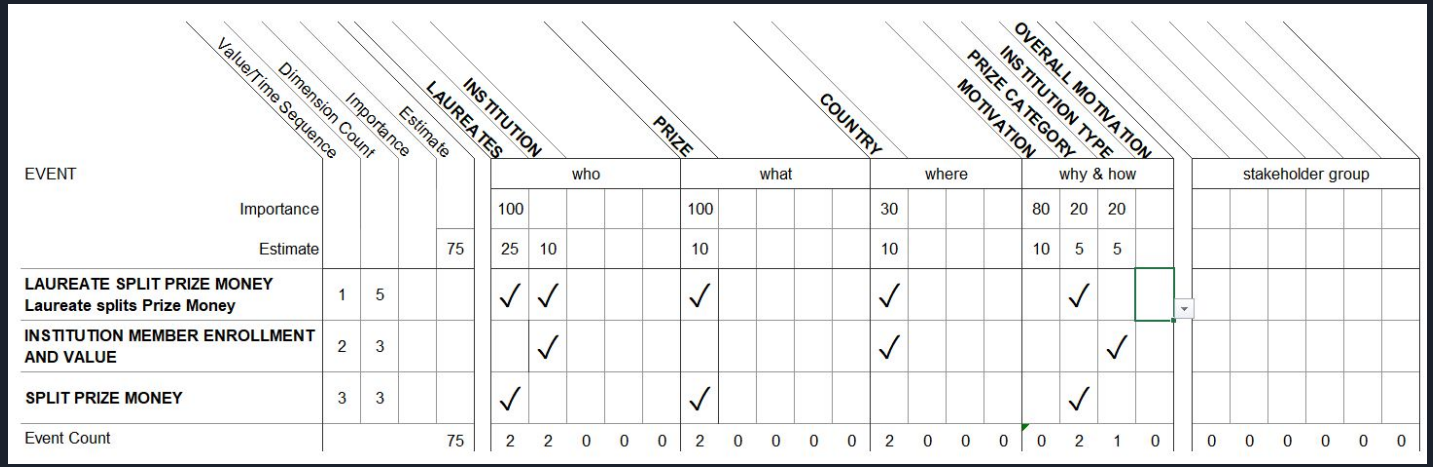
For the final project, our group chose to create a data warehouse to determine if a person's origin country correlates to their likelihood of winning a Nobel Prize. We would like to use our data warehouse to determine if the development of a country affects the amount of people born there who have won Nobel Prizes. For example, is it appropriate to theorize that a person from a first world country is more likely to win a Nobel Prize, compared to someone from a developing country?

Data Source

The data that we used comes from the Harvard Dataverse, which has published a dataset that contains three files; information about Nobel laureates, the prize they have won, and a list of the countries. We use this data to create dimensions such as laureate, overall motivation, prize category and institution.

Preparation and Planning

BEAM Matrix



EVENT	Value/Time Sequence	Dimension Count	Importance	Estimate	INSTITUTION	LAUREATES	PRIZE	COUNTRY	PRIZE CATEGORY	MOTIVATION	OVERALL MOTIVATION	INSTITUTION TYPE	stakeholder group
					who		what		where		why & how		
Importance					100		100		30		80	20	20
Estimate				75	25	10	10		10		10	5	5
LAUREATE SPLIT PRIZE MONEY Laureate splits Prize Money	1	5			✓	✓	✓		✓			✓	
INSTITUTION MEMBER ENROLLMENT AND VALUE	2	3				✓			✓				✓
SPLIT PRIZE MONEY	3	3			✓		✓					✓	
Event Count				75	2	2	2		2		0	2	1

The BEAM event matrix which shows the various dimensions and events.

BEAM Dimensions

LAUREATES	HV													
Laureate ID	First Name	Last Name	Date of Birth	Date of Death	Birth Country	Birth City	Gender	Year Prize Received	Prize Category	Prize Amount	Motivation	Share	Institution Name	Institution Country
BK, NN	{LAUREATE.CSV}	{LAUREATE.CSV}	FV {LAUREATE.CSV}	{LAUREATE.CSV}	FV {LAUREATE.CSV}	FV {LAUREATE.CSV}	{LAUREATE.CSV}	FV, NN {LAUREATE.CSV}	FV, NN {LAUREATE.CSV}	NN {LAUREATE.CSV}	{LAUREATE.CSV}	NN {LAUREATE.CSV}	{LAUREATE.CSV}	{LAUREATE.CSV}
I	C100	C100	D	D	C100	C100	C10	D	C20	N	C300	I	C150	C100
1	Wilhelm Conrad	Rantgen	1845-03-27	2/10/1923	Prussia (now Germany)	Lennep (now Remscheid)	male	1901	physics	1000000	"in recognition"	1	Munich University	Germany
2	Hendrik Antoon	Lorentz	1853-07-18	2/4/1928	the Netherlands	Arnhem	male	1902	physics	500000	"for discovery"	2	Leiden University	the Netherlands
3	Pieter	Zeeman	1865-05-25	10/9/1943	the Netherlands	Zonnemaire	male	1902	physics	500000	"for recognition"	2	Amsterdam University	the Netherlands
4	Antoine Henri	Becquerel	1852-12-15	8/25/1908	France	Paris	male	1903	physics	500000	"in recognition"	2	Polytechnique	France
5	Pierre	Curie	1859-05-15	4/19/1906	France	Paris	male	1903	physics	250000	"for discovery"	4	(Municipal School of Industrial Physics and Chemistry)	France
6	Marie	Curie Sklodowska	1867-11-07	7/4/1934	Russian Empire (now Poland)	Warsaw	female	1903	physics	250000	"for recognition"	4	Frankfurt-on-the-Main University	Germany
6	Marie	Curie, Sklodowska	1867-11-07	7/4/1934	Russian Empire (now Poland)	Warsaw	female	1911	chemistry	1000000	"in recognition"	1	Sorbonne University	France
8	Lord Rayleigh	John William Strutt	1842-11-12	6/30/1919	United Kingdom	Langford Grove, Maldon, Essex	male	1904	physics	1000000	"for discovery"	1	Royal Institution of Great Britain	United Kingdom
9	United Nations	NULL	0000-00-00	0000-00-00	NULL	NULL	NULL	1905	peace	1000000	"for recognition"	1	NULL	NULL

A screenshot of the Laureates dimension from the BEAM Modelstormer.

BEAM Dimensions

INSTITUTION

Institution Key	Institution ID	Institution Name	Institution City	Institution Country
SK, NN	BK, NN	NN, ND		
	{Institution.CSV}	{Institution.CSV}	{Institution.CSV}	{Institution.CSV}
I	C6	C100	C100	C100
1	IN001	Kiel University	Kiel	Germany
2	IN002	University of Cambridge	Cambridge	United Kingdom
3	IN003	University of Chicago	Chicago, IL	USA
4	IN004	Sorbonne University	Paris	France
5	IN005	Marconi Wireless Telegraph Co. Ltd.	London	United Kingdom
6	IN006	Strasbourg University	Strasbourg	Alsace (then Germany, now France)
7	IN007	Amsterdam University	Amsterdam	the Netherlands
8	IN008	Harvard	Cambridge, MA	United States
9	IN009	Swedish Gas-Accumulator Co.	Stockholm	Sweden

A screenshot of the Institution dimension from the BEAM Modelstormer.

BEAM Dimensions

INSTITUTION TYPE			
Institution Type Key	**Institution Type ID**	**Institution Type**	**Institution Type Description**
SK, NN **I**	**BK, NN** **{InstitutionType.CSV}**	**{InstitutionType.CSV}** **C30**	**{InstitutionType.CSV}** **C200**
1	INT001	Educational	involves schools and universities either being private or public
2	INT002	Religious	involves religious groups i.e. Christianity, Judaism, Islam, Hinduism, Buddhism, and Sikhism
3	INT003	Medicinal	involves hospitals and other health care institutions either being private or public
4	INT004	Government	involves governments that are either monarchy, oligarchy, dictatorship, and democracy
5	INT005	Economic	involves foundations focused on collecting economic data with the idea of providing a good or service that is important to the economy
6	INT006	Research	involves establishments that speciaize in advancing scholarly activity through research and experimental development

A screenshot of the Institution Type dimension from the BEAM Modelstormer.

BEAM Dimensions

COUNTRY			
Country Key	Country ID	Country Name	Country Abbreviation
SK, NN	BK, NN	NN	
	{Country.CSV}	{Country.CSV}	{Country.CSV}
I	C5	C100	C3
1	C100	Alsace, then Germany	DE
2	C101	Alsace	DE
3	C102	Germany	DE
4	C103	Argentina	AR
5	C104	Australia	AU
6	C105	Austria	AT
7	C106	Belgium	BE
8	C107	Burma	MM
9	C108	Old Republic	NULL

A screenshot of the Country dimension from the BEAM Modelstormer.

BEAM Dimensions

PRIZE			
Prize Key	Prize ID	Prize Year	Prize Category
SK, NN	BK, NN	NN	NN
	Prize.CSV}	{Prize.CSV}	{Prize.CSV}
I	C5	C4	C20
1	P001	2016	physics
2	P002	2016	physics
3	P003	2016	physics
4	P004	2016	chemistry
5	P005	2016	chemistry
6	P006	2016	chemistry
7	P007	2016	medicine
8	P008	2016	literature
9	P009	2016	peace

PRIZE CATEGORY		
Category Key	Category ID	Category
SK, NN	BK, NN	ND
	{Category.CSV}	{Category.CSV}
I	C5	C20
1	NPC1	physics
2	NPC2	chemistry
3	NPC3	medicine
4	NPC4	peace
5	NPC5	literature
6	NPC6	economics

A screenshot of the Prize and Prize Category from the BEAM Modelstormer.

BEAM Dimensions

MOTIVATION		
Motivation Key	Motivation ID	Motivation
SK, NN	BK, NN	NN
	{Motivation.CSV}	{Motivation.CSV}
I	C5	C300
1	M001	in recognition of his work in thermochemistry
2	M002	for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material
3	M003	for their discoveries concerning the structural and functional organization of the cell
4	M004	for the discovery of the quantized Hall effect
5	M005	for their pathbreaking contribution to the theory of international trade and international capital movements
6	M006	for having created new poetic expressions within the great American song tradition
7	M007	for his discovery of the organizer effect in embryonic development
8	M008	for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution
9	M009	for their discoveries relating to the hormones of the adrenal cortex, their structure and biological effects

OVERALL MOTIVATION		
Overall Motivation Key	Overall Motivation ID	Overall Motivation
SK, NN	BK, NN	NN
	{OverallMotivation.CSV}	{OverallMotivation.CSV}
I	C5	C100
1	OM001	for pioneering experimental contributions to lepton physics
2	OM002	for contributions to the developments of methods within DNA-based chemistry
3	OM003	for basic work on information and communication technology
4	OM004	for the development of methods for identification and structure analyses of biological macromolecules
5	OM005	for discoveries concerning channels in cell membranes
6	OM006	for pioneering contributions to the development of neutron scattering techniques for studies of condensed matter

BEAM Facts

LAUREATE PRIZE						
LAUREATE KEY	originates COUNTRY KEY	belongs INSTITUTION KEY	receives PRIZE KEY	in CATEGORY KEY	SHARE	earned PRIZE AWARD
[who]	[where]	[what]	[what]	[what]	[how]	[$]
53	87	88	124	2	1	1000000
88	81	86	11	3	2	500000
30	17	47	7	2	2	500000
5	70	89	13	1	2	500000
59	21	75	134	2	4	250000
26	47	14	37	1	4	250000
84	22	80	76	2	1	1000000
91	87	4	188	6	1	1000000
73	37	77	196	6	1	1000000

A screenshot of the Laureate Shared Prize fact from the BEAM Modelstormer.

BEAM Facts

INSTITUTION VALUE & MEMBERS

Institution Key	Institution Type Key	Country Key	has Institution Value in Millions	with Institution Member Count
[who/what]	[what]	[where]	[$]	
88	1	87	16.6	424
86	2	81	723.4	7744
47	3	17	40.3	399333
89	4	70	51.7	234
75	5	21	3.2	32
14	6	47	1.9	6533
80	1	22	500	66250
4	2	87	221.2	29309
77	3	37	3.3	437

A screenshot of the Institution Value and Members fact from the BEAM Modelstormer.

BEAM Facts

INSTITUTION VALUE & MEMBERS

Institution Key	Institution Type Key	Country Key	has Institution Value in Millions	with Institution Member Count
[who/what]	[what]	[where]	[$]	
88	1	87	16.6	424
86	2	81	723.4	7744
47	3	17	40.3	399333
89	4	70	51.7	234
75	5	21	3.2	32
14	6	47	1.9	6533
80	1	22	500	66250
4	2	87	221.2	29309
77	3	37	3.3	437

A screenshot of the Institution Value and Members fact from the BEAM Modelstormer.

BEAM Facts

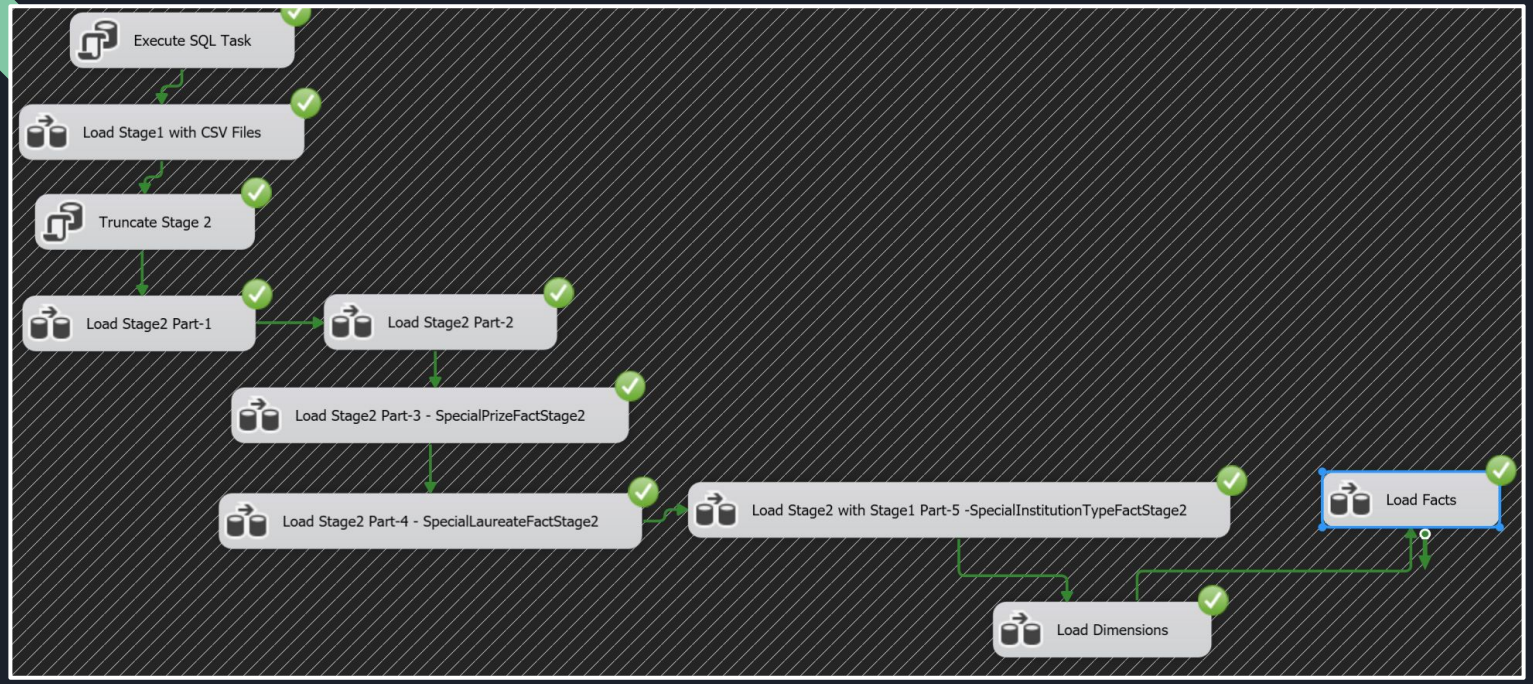
DISTRIBUTED MONEY

Laureate Key	recieves Prize Key	in Category Key	has to Share	results Prize Distribution per Winner
[who]	[what]		[how]	[$]
53	124	2	1	500000
88	11	3	2	250000
30	7	2	2	250000
5	13	1	2	333333.3333
59	134	2	4	333333.3333
26	37	1	4	333333.3333
84	76	2	1	1000000
91	188	6	1	1000000
73	196	6	1	1000000

A screenshot of the Distributed Money fact from the BEAM Modelstormer.

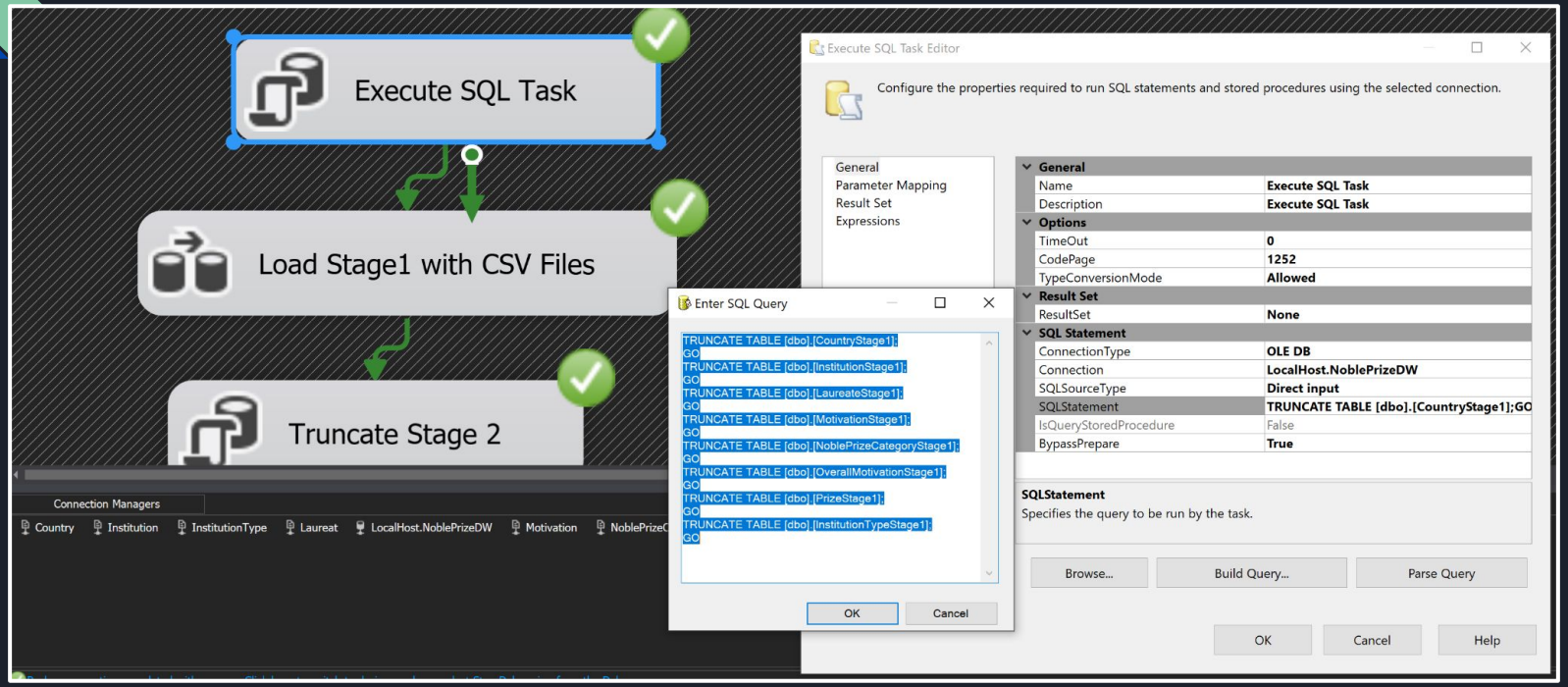
ETL Process

ETL Process: Whole Diagram



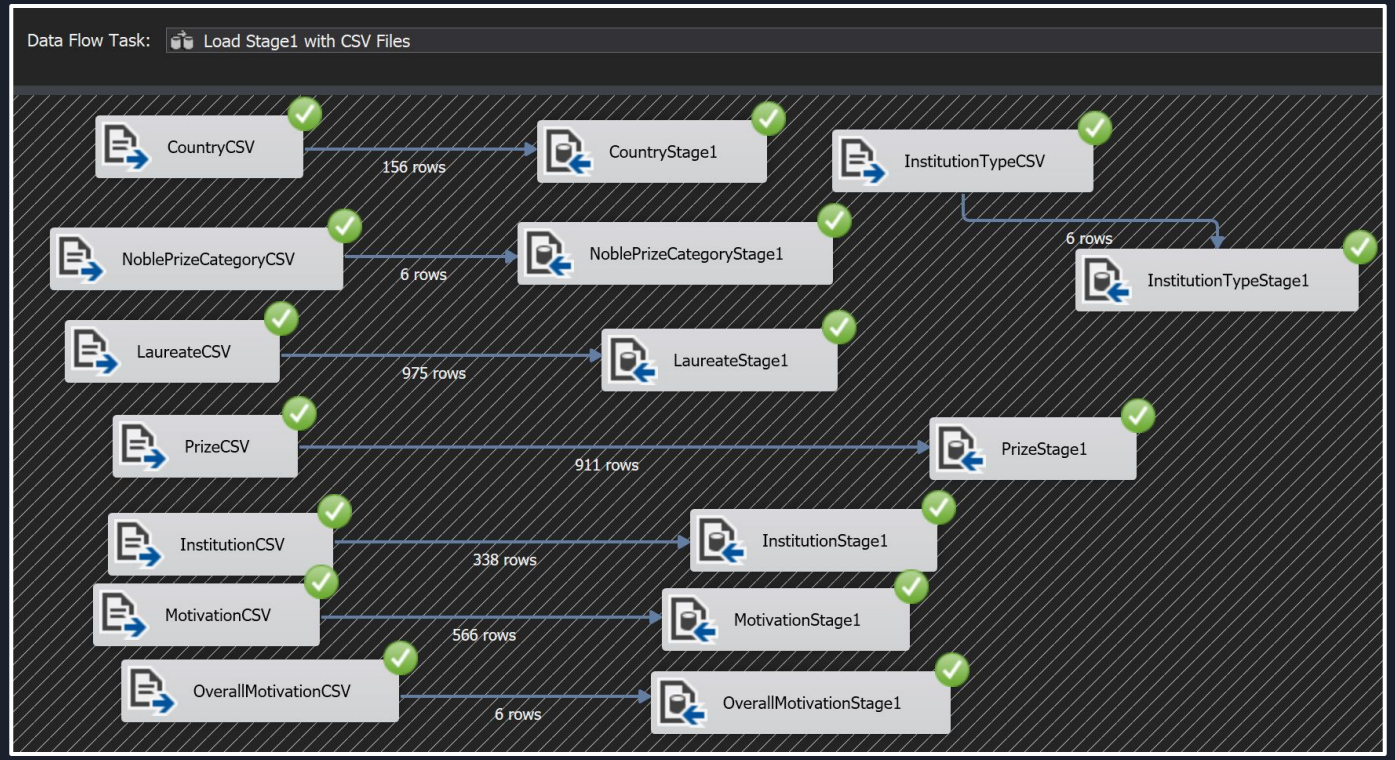
Shows the entire ETL Process' Steps from "Execute SQL Task" to "Load Facts"

ETL Process: Execute SQL Task



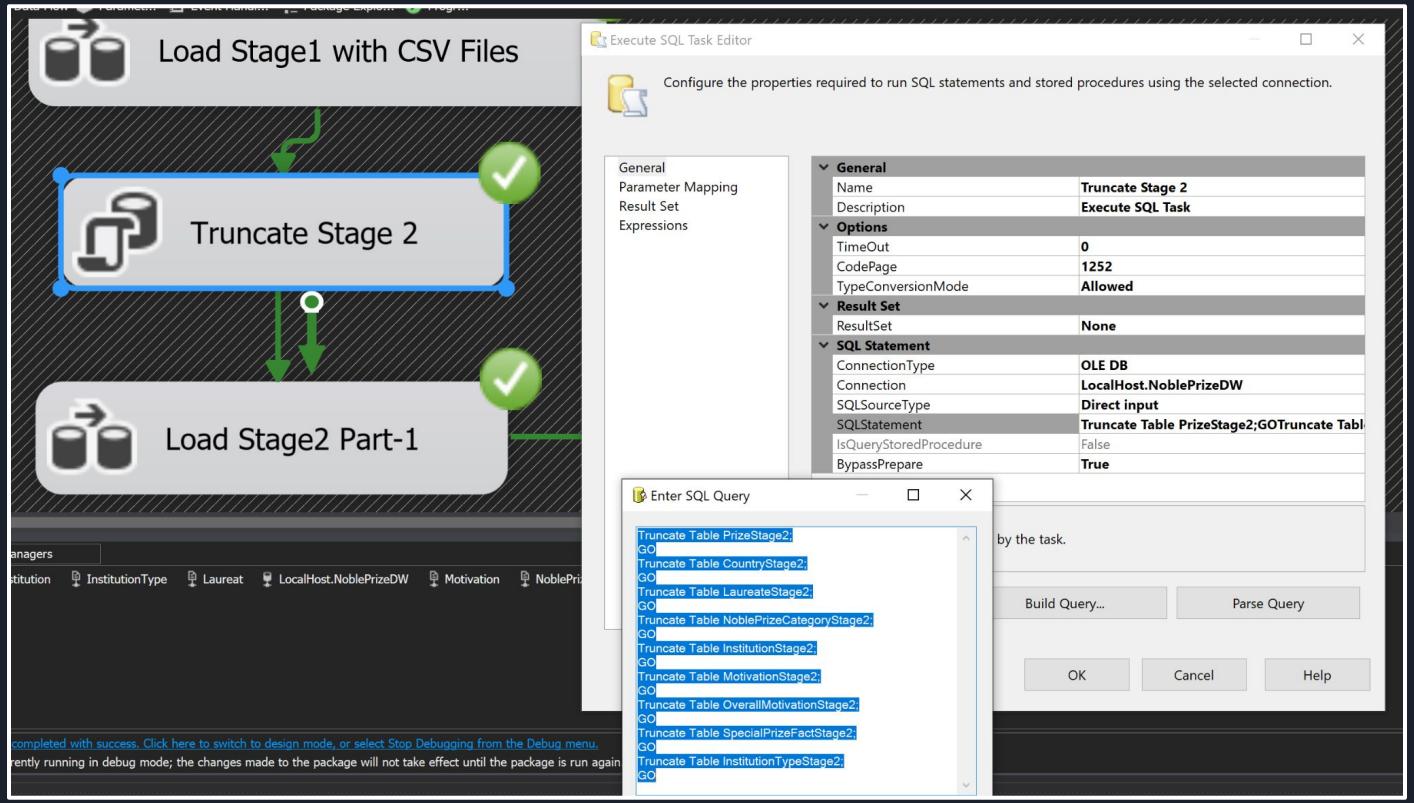
Extracts the data from the 8 CSV files to Load Stage1 Data Flow Task through Truncation

ETL Process: Load Stage1 with CSV Files



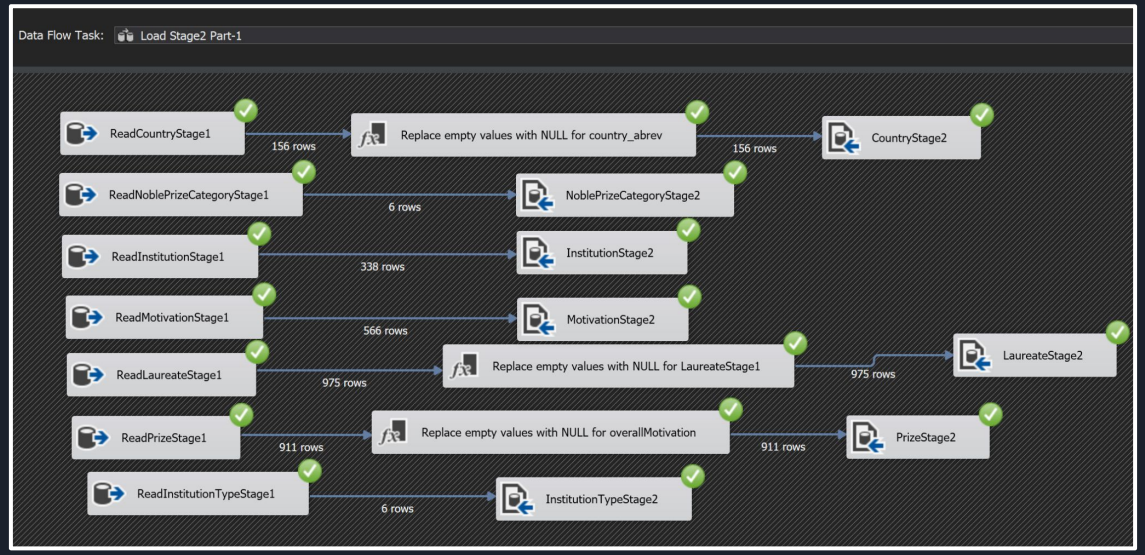
Extracts the data from the 8 CSV files to Stage1 SQL Server Destinations

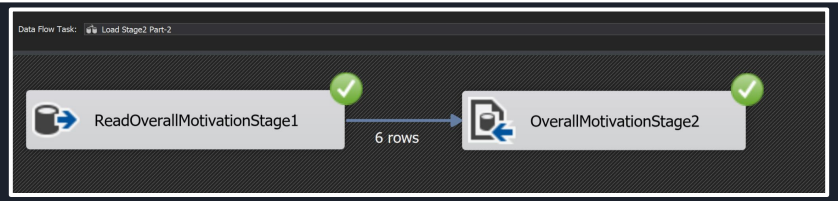
ETL Process: Truncate Stage 2



Extracts the data from the the Stage1 to Load Stage 2 Data Flow Task through Truncation

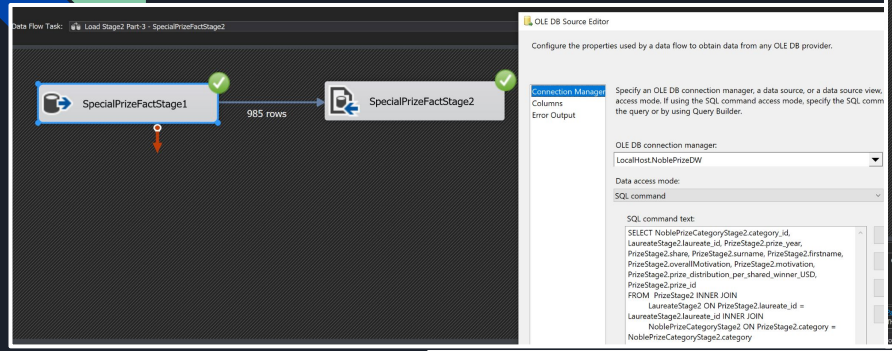
ETL Process: Load Stage2 with Stage1 Part 1 and 2

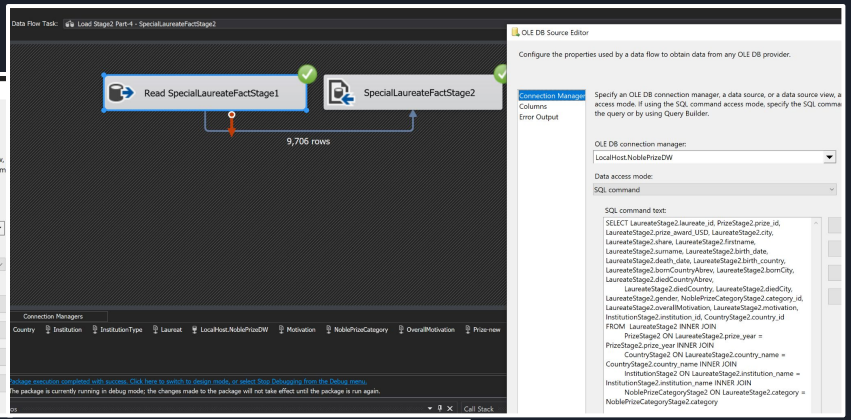


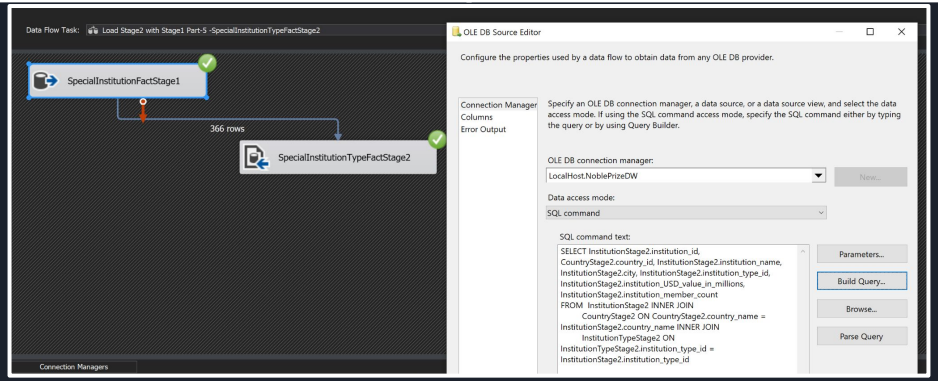


Extracts the data from the Stage1 OLE DB Sources to Stage 2 SQL Server Destinations with some Stages Replacing the Empty Values with NULL

ETL Process: Load Stage2 with Stage1 Part 3,4, and 5

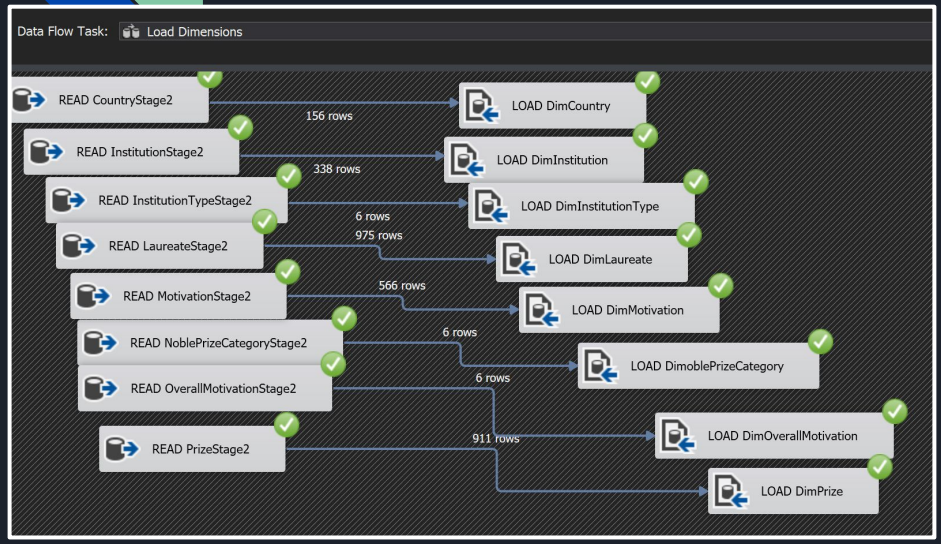


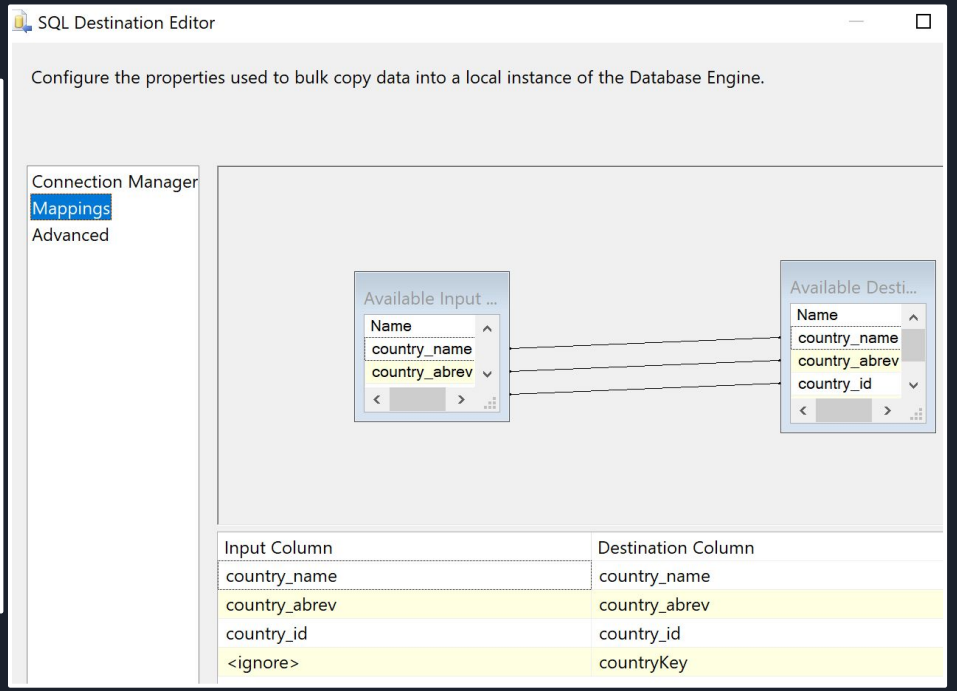




Extracts the data from the Stage1 OLE DB Sources to Stage 2 SQL Server Destinations with Stage 1 made through a SQL Commands

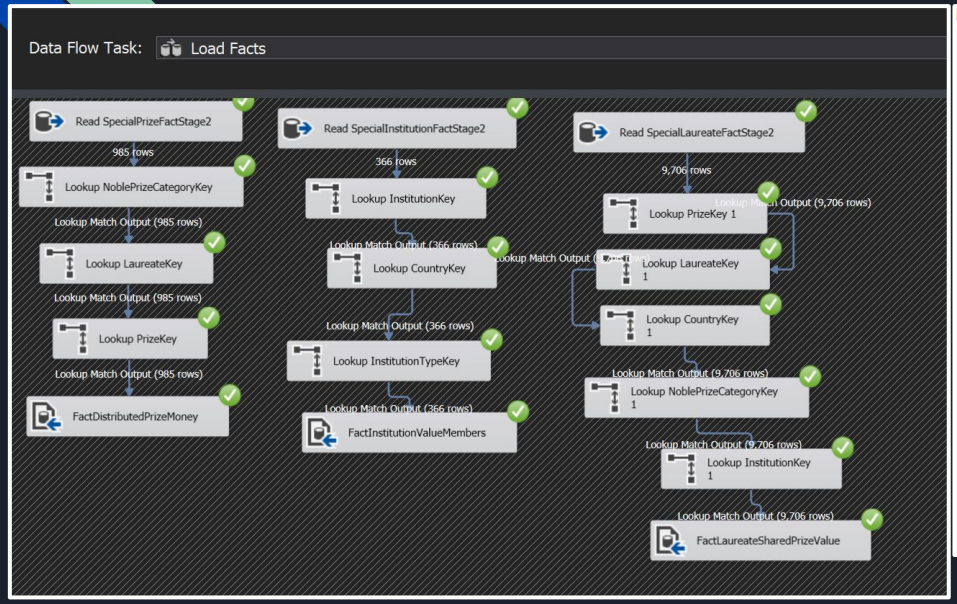
ETL Process: Load Dimensions

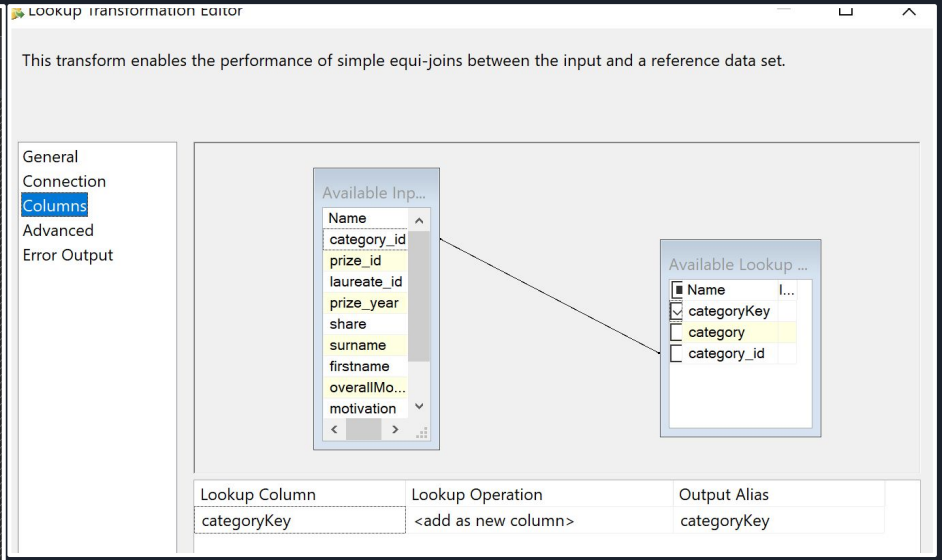




Extracts the data from the Stage2 OLE DB Sources to Load Dimension SQL Server Destinations with Loading Dimension SQL Server Destinations Creating Surrogate Keys that Replaces the Natural Keys of each Stage

ETL Process: Load Facts - Part One

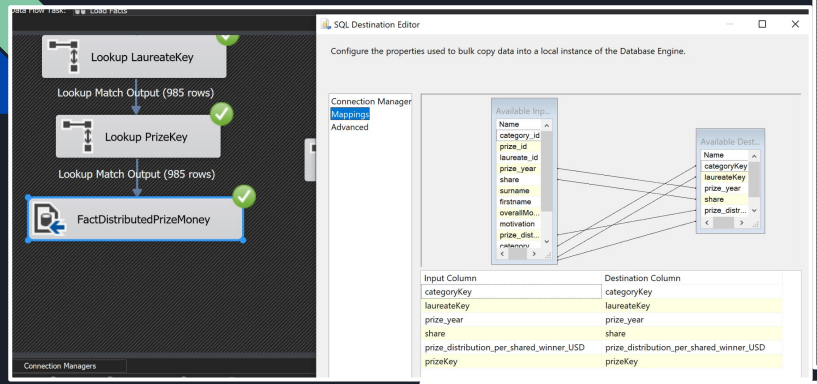


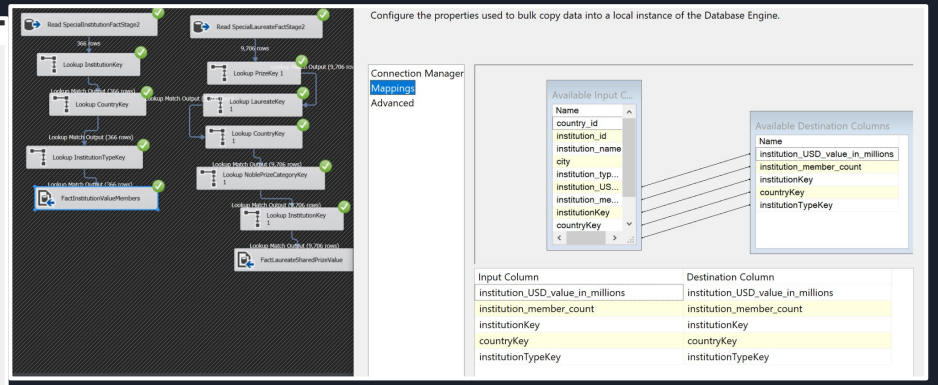


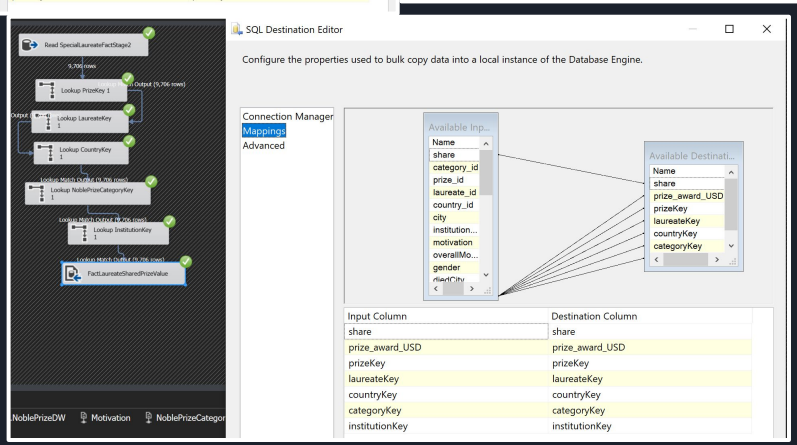
Extracts Data from Load Fact Stage 2 OLE DB Sources to Fact SQL Server Destination with Lookup Surrogate Key Pipelines for each Fact

Lookup Surrogate Key Pipeline of Dimension Category as Example

ETL Process: Load Facts - Part Two







Extracts the data from the Stage2 OLE DB Sources to Fact SQL Server Destinations with Lookup Surrogate Key Pipelines

Data Warehouse Mart Views

Nobel Prize Data Warehouse with 3 Marts

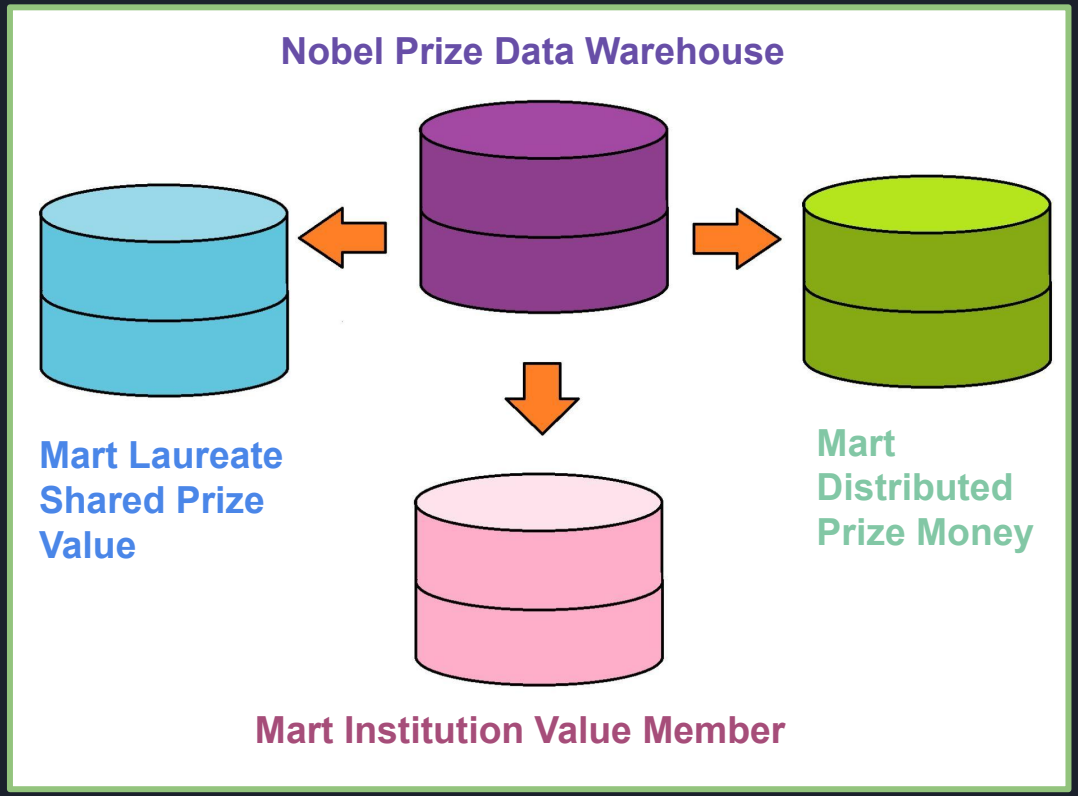
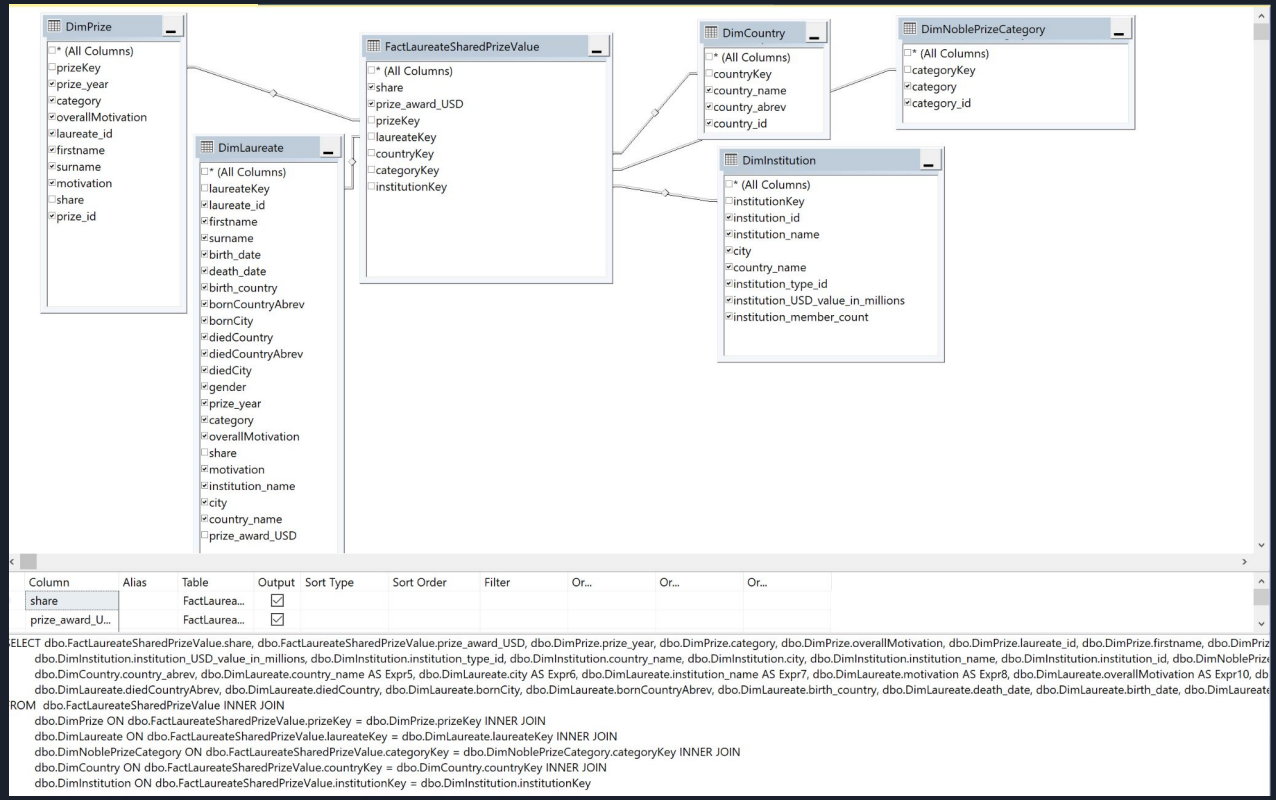


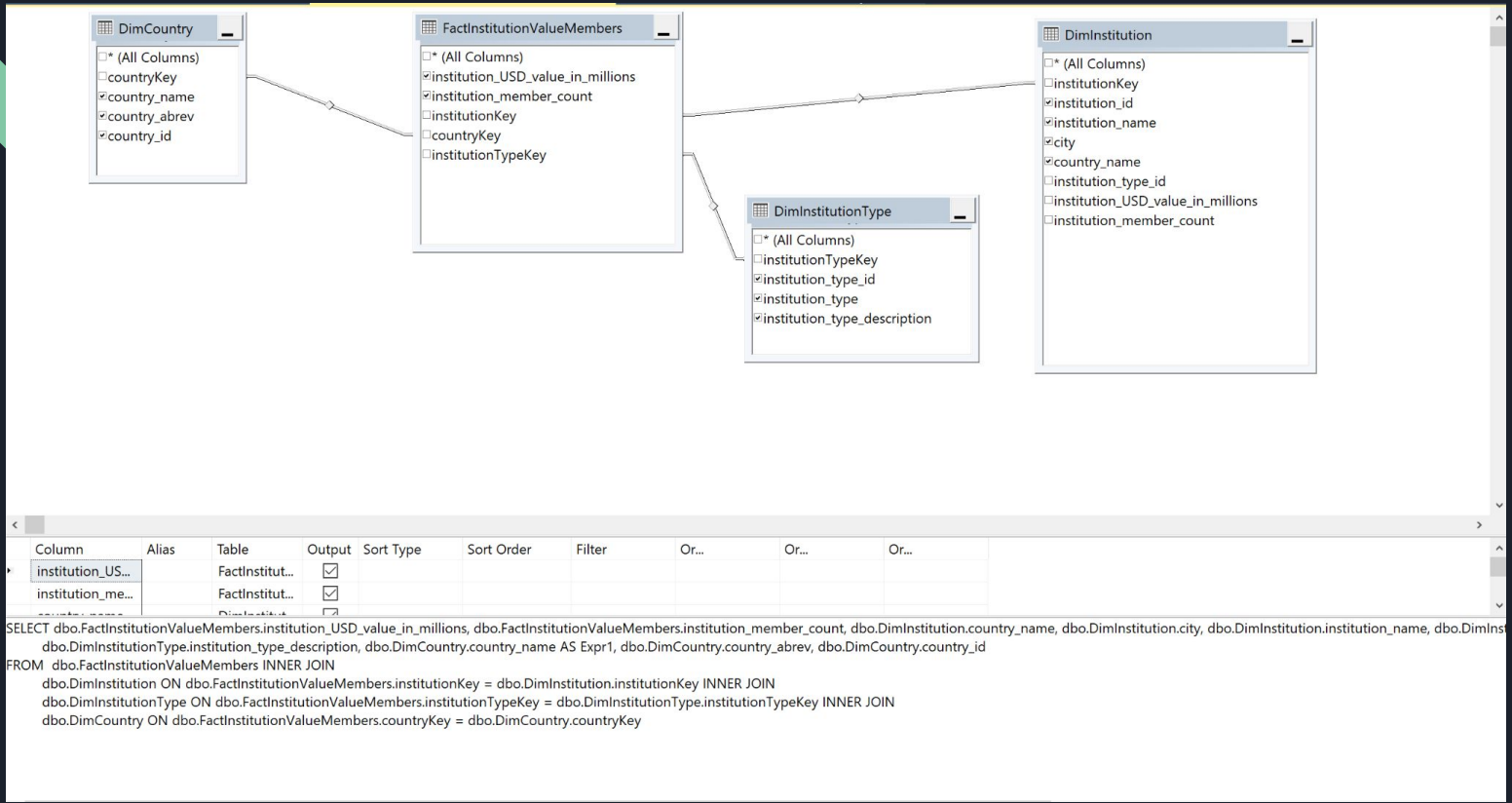
Diagram of the Nobel Prize Data Warehouse with the Laureate Shared Prize Value Mart, Institution Value Member Mart, and Distributed Prize Money Mart

Mart Laureate Shared Prize Value



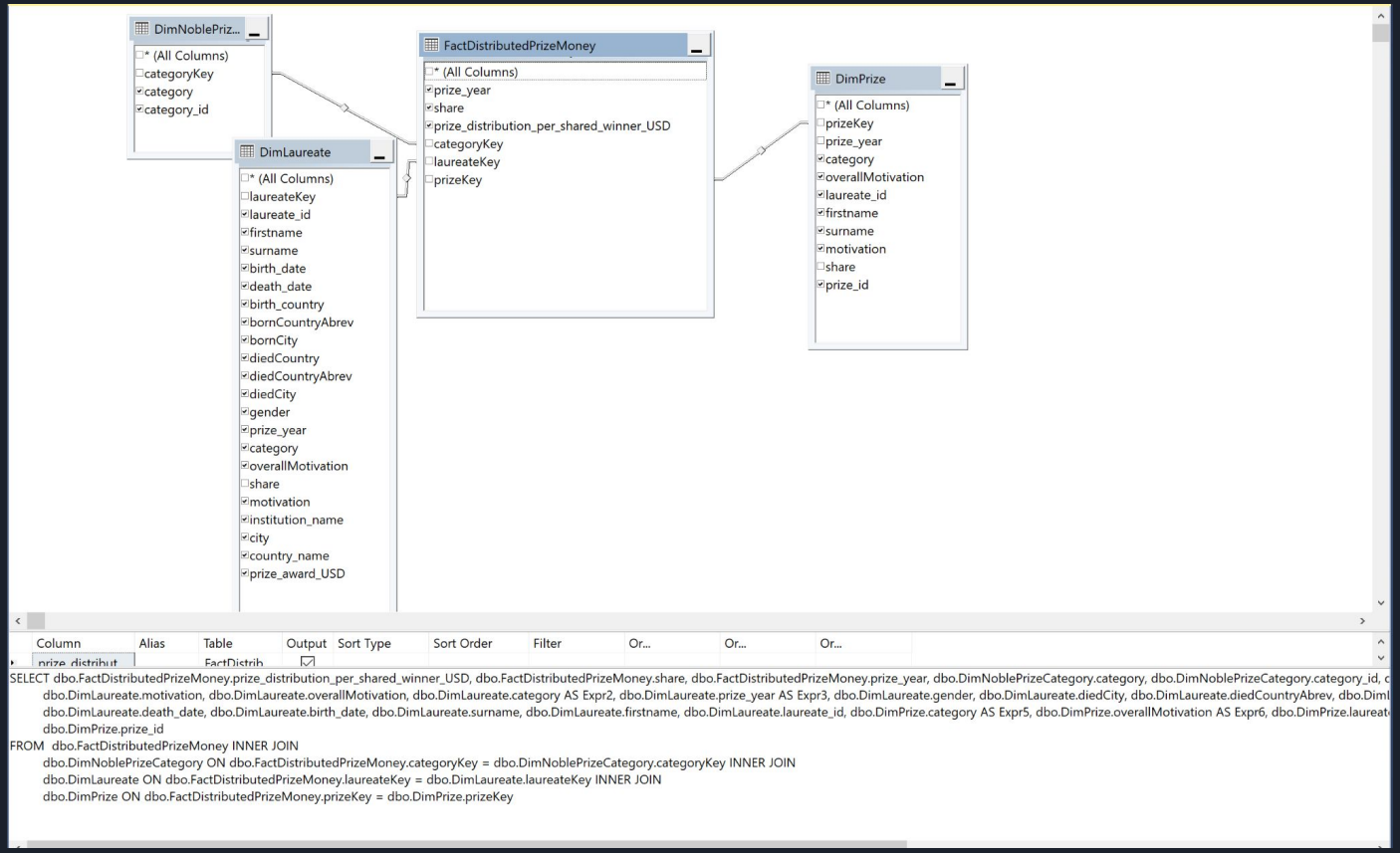
Nobel Prize Data Warehouse View of the Laureate Shared Prize Value Mart with the Laureate Shared Prize Value Fact Table and the Prize, Laureate, Country, Institution, and Noble Prize Category Dimension Tables

Mart Institution Value Member



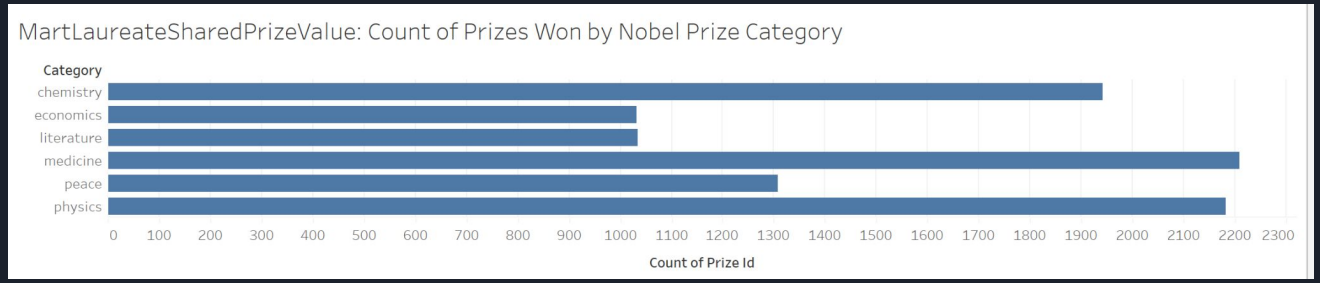
Nobel Prize Data Warehouse View of the Institution Value Members Mart with the Institution Value Members Fact Table and the Country, Institution Type, and Institution Dimension Tables

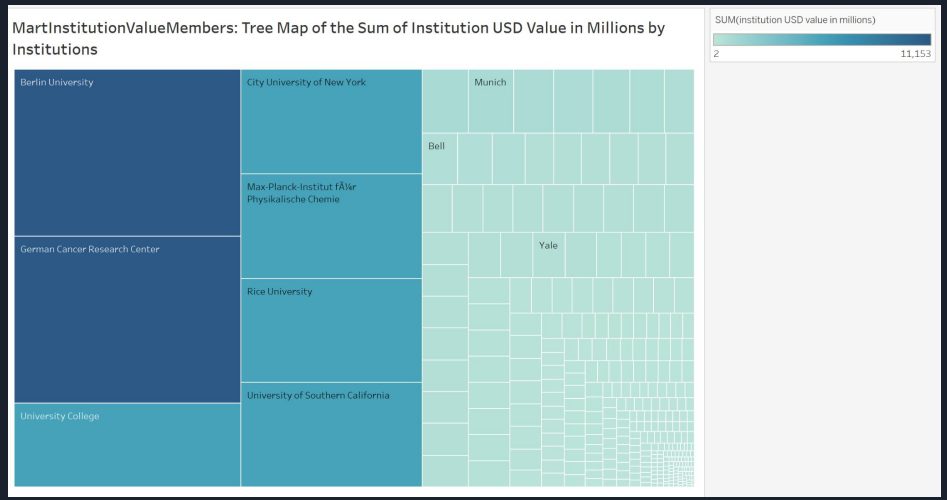
Mart Distributed Prize Money

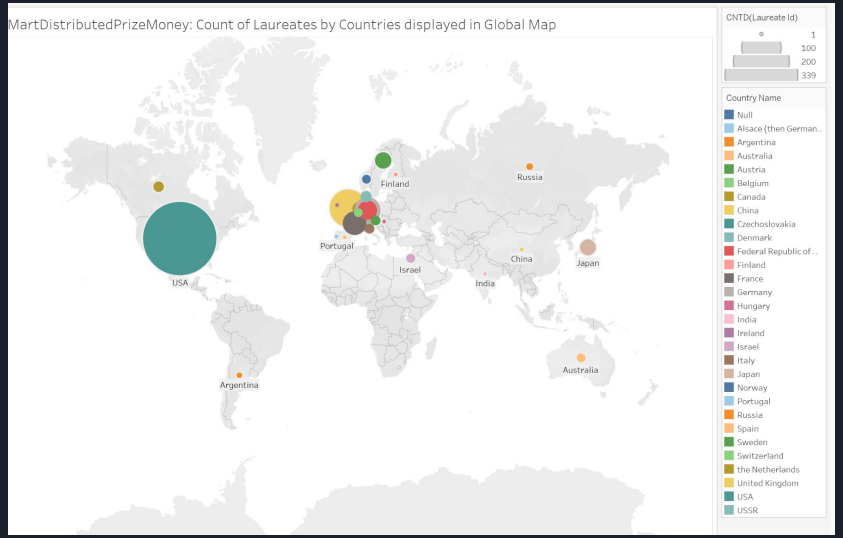


Nobel Prize Data Warehouse View of the Distribution Prize Money Mart with the Distributed Prize Money Fact Table and the Laureate, Nobel Prize Categories, and Prize Dimension Tables

Tableau Data Visualizations







SQL Queries

```sql
SELECT [prize_year], [1] as SHARED_ONE, [2] as SHARED_TWO, [3] as SHARED_THREE, [4] as SHARED_FOUR
FROM
(
    SELECT [prize_year], [share], [prize_distribution_per_shared_winner_USD]
    FROM [NoblePrizeDW].[dbo].[MartDistributedPrizeMoney]
) ps
PIVOT
(   AVG([prize_distribution_per_shared_winner_USD])
    FOR [share] IN ([1],[2],[3],[4])
)AS pvt
```

Results | Messages

	prize_year	SHARED_ONE	SHARED_TWO	SHARED_THREE	SHARED_FOUR
1	1901	1000000.000000	500000.000000	NULL	NULL
2	1902	1000000.000000	500000.000000	NULL	NULL
3	1903	1000000.000000	500000.000000	NULL	250000.000000
4	1904	1000000.000000	500000.000000	NULL	NULL
5	1905	1000000.000000	NULL	NULL	NULL
6	1906	1000000.000000	500000.000000	NULL	NULL
7	1907	1000000.000000	500000.000000	NULL	NULL
8	1908	1000000.000000	500000.000000	NULL	NULL
9	1909	1000000.000000	500000.000000	NULL	NULL
10	1910	1000000.000000	NULL	NULL	NULL
11	1911	1000000.000000	500000.000000	NULL	NULL
12	1912	1000000.000000	500000.000000	NULL	NULL
13	1913	1000000.000000	NULL	NULL	NULL
14	1914	1000000.000000	NULL	NULL	NULL
15	1915	1000000.000000	500000.000000	NULL	NULL
16	1916	1000000.000000	NULL	NULL	NULL
17	1917	1000000.000000	500000.000000	NULL	NULL
18	1918	1000000.000000	NULL	NULL	NULL
19	1919	1000000.000000	NULL	NULL	NULL
20	1920	1000000.000000	NULL	NULL	NULL
21	1921	1000000.000000	500000.000000	NULL	NULL
22	1922	1000000.000000	500000.000000	NULL	NULL
23	1923	1000000.000000	500000.000000	NULL	NULL
24	1924	1000000.000000	NULL	NULL	NULL
25	1925	1000000.000000	500000.000000	NULL	NULL
26	1926	1000000.000000	500000.000000	NULL	NULL
27	1927	1000000.000000	500000.000000	NULL	NULL

```sql
SELECT InstitutionStage2.institution_id, CountryStage2.country_id, InstitutionStage2.institution_name,
InstitutionStage2.city, InstitutionStage2.institution_type_id, InstitutionStage2.institution_USD_value_in_millions,
InstitutionStage2.institution_member_count
FROM  [NoblePrizeDW].[dbo].[InstitutionStage2 INNER JOIN
        [NoblePrizeDW].[dbo].CountryStage2 ON CountryStage2.country_name = InstitutionStage2.country_name INNER JOIN
       [NoblePrizeDW].[dbo].InstitutionTypeStage2 ON
       InstitutionTypeStage2.institution_type_id = InstitutionStage2.institution_type_id
```

Results | Messages

institution_id	country_id	institution_name	city	institution_type_id	institution_USD_value_in_millions	institution_member_count
IN001	C142	A.F. Ioffe Physico-Technical Institute	St. Petersburg	INT001	16.6	424
IN002	C121	Ã¼cole municipale de physique et de chimie indu...	Paris	INT001	723.4	7744
IN003	C121	Ã¼cole Nationale SupÃ©rieur des Mines de Paris	Paris	INT002	40.3	399333
IN004	C121	Ã¼cole Normale SupÃ©rieure	Paris	INT004	51.7	234
IN005	C121	Ã¼cole Polytechnique	Paris	INT005	3.2	32
IN006	C121	Ã¼cole SupÃ©rieure de Physique et Chimie	Paris	INT006	1.9	6533
IN007	C116	Aarhus University	Aarhus	INT002	500	66250
IN008	C154	Academy of Sciences	Moscow	INT003	221.2	29309
IN009	C124	All Souls College	Oxford	INT001	3.3	437
IN010	C149	Amsterdam University	Amsterdam	INT004	77.4	10255
IN011	C153	Argonne National Laboratory	Argonne, IL	INT005	428.6	56789
IN012	C153	Arizona State University	Tempe, AZ	INT006	22.3	2954
IN013	C153	Associated Universities Inc.	Washington, DC	INT003	7.32	969
IN014	C124	Atomic Energy Research Establishment	Harwell, Berkshire	INT004	7.564	1002
IN015	C104	Australian National University	Canberra	INT005	90	11925
IN016	C104	Australian National University	Weston Creek	INT001	4.6	609
IN017	C148	Basel Institute for Immunology	Basel	INT002	4.5	596
IN018	C148	Basel University	Basel	INT004	42.4	5618
IN019	C153	Bell Laboratories	Holmdel, NJ	INT005	323.5	42863
IN020	C153	Bell Laboratories	Murray Hill, NJ	INT006	210.4	27878
IN021	C153	Bell Telephone Laboratories	Murray Hill, NJ	INT003	32.3	4279
IN022	C153	Bell Telephone Laboratories	New York, NY	INT001	22.3	2954
IN023	C102	Berlin University	Berlin	INT002	5576.3	738859
IN023	C122	Berlin University	Berlin	INT002	5576.3	738859
IN024	C148	Berne University	Berne	INT003	95.4	12640
IN025	C148	Biozentrum der UniversitÃ¤t	Basel	INT001	33.4	4425
IN026	C124	Birmingham University	Birmingham	INT002	18.4	2438

```sql
SELECT [NoblePrizeDW].[dbo].[NoblePrizeCategoryStage2].[category_id],
[NoblePrizeDW].[dbo].[LaureateStage2].[laureate_id], PrizeStage2.prize_year, PrizeStage2.share,
PrizeStage2.surname, PrizeStage2.firstname, PrizeStage2.overallMotivation, PrizeStage2.motivation,
PrizeStage2.prize_distribution_per_shared_winner_USD, PrizeStage2.prize_id
FROM  [NoblePrizeDW].[dbo].PrizeStage2 INNER JOIN
        [NoblePrizeDW].[dbo].[LaureateStage2] ON [NoblePrizeDW].[dbo].PrizeStage2.laureate_id = [NoblePrizeDW].[dbo].LaureateStage2.laureate_id
        INNER JOIN [NoblePrizeDW].[dbo].NoblePrizeCategoryStage2 ON
        [NoblePrizeDW].[dbo].PrizeStage2.category = [NoblePrizeDW].[dbo].NoblePrizeCategoryStage2.category
```

Results | Messages

category_id	laureate_id	prize_year	share	surname	firstname	overallMotivation	motivation	prize_distribution_per_shared_winner_USD	prize_id
NPC1	1	1901	1	Röntgen	Wilhelm Conrad	NULL	"in recognition of the extraordinary services he ha...	1000000	P906
NPC1	10	1906	1	Thomson	Joseph John	NULL	"in recognition of the great merits of his theoretical...	1000000	P876
NPC1	100	1974	2	Ryle	Sir Martin	NULL	"for their pioneering research in radio astrophysics:...	500000	P481
NPC1	101	1974	2	Hewish	Antony	NULL	"for their pioneering research in radio astrophysics:...	500000	P482
NPC1	102	1975	3	Bohr	Aage Niels	NULL	"for the discovery of the connection between coll...	333333.3333	P449
NPC1	103	1975	3	Mottelson	Ben Roy	NULL	"for the discovery of the connection between coll...	333333.3333	P450
NPC1	104	1975	3	Rainwater	Leo James	NULL	"for the discovery of the connection between coll...	333333.3333	P451
NPC1	105	1976	2	Richter	Burton	NULL	"for their pioneering work in the discovery of a he...	500000	P440
NPC1	106	1976	2	Ting	Samuel Chao Chung	NULL	"for their pioneering work in the discovery of a he...	500000	P441
NPC1	107	1977	3	Anderson	Philip Warren	NULL	"for their fundamental theoretical investigations of t...	333333.3333	P429
NPC1	108	1977	3	Mott	Sir Nevill Francis	NULL	"for their fundamental theoretical investigations of t...	333333.3333	P430
NPC1	109	1977	3	van Vleck	John Hasbrouck	NULL	"for their fundamental theoretical investigations of t...	333333.3333	P431
NPC1	11	1907	1	Michelson	Albert Abraham	NULL	"for his optical precision instruments and the spec...	1000000	P869
NPC1	110	1978	2	Kapitsa	Pyotr Leonidovich	NULL	"for his basic inventions and discoveries in the are...	500000	P418
NPC1	111	1978	4	Penzias	Arno Allan	NULL	"for their discovery of cosmic microwave backgro...	250000	P419
NPC1	112	1978	4	Wilson	Robert Woodrow	NULL	"for their discovery of cosmic microwave backgro...	250000	P420
NPC1	113	1979	3	Glashow	Sheldon Lee	NULL	"for their contributions to the theory of the unified ...	333333.3333	P427
NPC1	114	1979	3	Salam	Abdus	NULL	"for their contributions to the theory of the unified ...	333333.3333	P428
NPC1	114	1979	3	Salam	Abdus	NULL	"for their contributions to the theory of the unified ...	333333.3333	P408
NPC1	115	1979	3	Weinberg	Steven	NULL	"for their contributions to the theory of the unified ...	333333.3333	P409
NPC1	116	1980	2	Cronin	James Watson	NULL	"for the discovery of violations of fundamental sym...	500000	P396
NPC1	117	1980	2	Fitch	Val Logsdon	NULL	"for the discovery of violations of fundamental sym...	500000	P397
NPC1	118	1981	4	Bloembergen	Nicolaas	NULL	"for their contribution to the development of laser ...	250000	P385
NPC1	119	1981	4	Schawlow	Arthur Leonard	NULL	"for their contribution to the development of laser ...	250000	P386
NPC1	12	1908	1	Lippmann	Gabriel	NULL	"for his method of reproducing colours photograp...	1000000	P862
NPC1	120	1981	2	Siegbahn	Kai M.	NULL	"for his contribution to the development of high-re...	500000	P387
NPC1	121	1982	1	Wilson	Kenneth G.	NULL	"for his theory for critical phenomena in connectio...	1000000	P376

```sql
SELECT [NoblePrizeDW].[dbo].[LaureateStage2].laureate_id, [NoblePrizeDW].[dbo].[PrizeStage2].prize_id,
       [NoblePrizeDW].[dbo].[LaureateStage2].prize_award_USD, [NoblePrizeDW].[dbo].[LaureateStage2].city,
       [NoblePrizeDW].[dbo].[LaureateStage2].share, [NoblePrizeDW].[dbo].[LaureateStage2].firstname,
       [NoblePrizeDW].[dbo].[LaureateStage2].surname, [NoblePrizeDW].[dbo].[LaureateStage2].birth_date,
       [NoblePrizeDW].[dbo].[LaureateStage2].death_date, [NoblePrizeDW].[dbo].[LaureateStage2].birth_country,
       [NoblePrizeDW].[dbo].[LaureateStage2].bornCountryAbrev, [NoblePrizeDW].[dbo].[LaureateStage2].bornCity,
       [NoblePrizeDW].[dbo].[LaureateStage2].diedCountryAbrev, [NoblePrizeDW].[dbo].[LaureateStage2].diedCountry,
       [NoblePrizeDW].[dbo].[LaureateStage2].diedCity, [NoblePrizeDW].[dbo].[LaureateStage2].gender,
       [NoblePrizeDW].[dbo].[NoblePrizeCategoryStage2].category_id, [NoblePrizeDW].[dbo].[LaureateStage2].overallMotivation,
       [NoblePrizeDW].[dbo].[LaureateStage2].motivation, [NoblePrizeDW].[dbo].[InstitutionStage2].institution_id,
       [NoblePrizeDW].[dbo].[CountryStage2].country_id
FROM  [NoblePrizeDW].[dbo].[LaureateStage2] INNER JOIN
        [NoblePrizeDW].[dbo].[PrizeStage2] ON
        [NoblePrizeDW].[dbo].[LaureateStage2].prize_year = [NoblePrizeDW].[dbo].[PrizeStage2].prize_year INNER JOIN
        [NoblePrizeDW].[dbo].[CountryStage2] ON
        [NoblePrizeDW].[dbo].[LaureateStage2].country_name = [NoblePrizeDW].[dbo].[CountryStage2].country_name INNER JOIN
        [NoblePrizeDW].[dbo].[InstitutionStage2] ON
        [NoblePrizeDW].[dbo].[LaureateStage2].institution_name = [NoblePrizeDW].[dbo].[InstitutionStage2].institution_name INNER JOIN
        [NoblePrizeDW].[dbo].[NoblePrizeCategoryStage2] ON
        [NoblePrizeDW].[dbo].[LaureateStage2].category = [NoblePrizeDW].[dbo].[NoblePrizeCategoryStage2].category
```

Results | Messages

laureate_id	prize_id	prize_award_USD	city	share	firstname	surname	birth_date	death_date	birth_country	bornCountryAbrev	bornCity	diedCountryAbrev	diedCountry	diedCity	gender	category_id
927	P001	1000000	Tokyo	2	Yoshinori	Ohsumi	0000-00-00	0000-00-00	Japan	JP	Fukuoka	NULL	NULL	NULL	male	NPC3
928	P001	500000	Seattle, WA	2	David J.	Thouless	9/21/1934	0000-00-00	United Kingdom	GB	Bearsden	NULL	NULL	NULL	male	NPC1
929	P001	250000	Princeton, NJ	4	F. Duncan M.	Haldane	9/14/1951	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC1
930	P001	250000	Providence, RI	4	J. Michael	Kosterlitz	6/22/1942	0000-00-00	United Kingdom	GB	Aberdeen	NULL	NULL	NULL	male	NPC1
931	P001	333333.3333	Strasbourg	3	Jean-Pierre	Sauvage	10/21/1944	0000-00-00	France	FR	Paris	NULL	NULL	NULL	male	NPC2
932	P001	333333.3333	Evanston, IL	3	Sir J. Fraser	Stoddart	5/24/1942	0000-00-00	United Kingdom	GB	Edinburgh	NULL	NULL	NULL	male	NPC2
933	P001	333333.3333	Groningen	3	Bernard L.	Feringa	5/18/1951	0000-00-00	the Netherlands	NL	Barger-Compascuum	NULL	NULL	NULL	male	NPC2
934	P001	500000	Cambridge, MA	2	Oliver	Hart	10/9/1948	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC1
935	P001	500000	Cambridge, MA	2	Bengt	Holmström	4/18/1949	0000-00-00	Finland	FI	Helsinki	NULL	NULL	NULL	male	NPC6
927	P002	1000000	Tokyo	1	Yoshinori	Ohsumi	0000-00-00	0000-00-00	Japan	JP	Fukuoka	NULL	NULL	NULL	male	NPC3
928	P002	500000	Seattle, WA	2	David J.	Thouless	9/21/1934	0000-00-00	United Kingdom	GB	Bearsden	NULL	NULL	NULL	male	NPC1
929	P002	250000	Princeton, NJ	4	F. Duncan M.	Haldane	9/14/1951	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC1
930	P002	250000	Providence, RI	4	J. Michael	Kosterlitz	6/22/1942	0000-00-00	United Kingdom	GB	Aberdeen	NULL	NULL	NULL	male	NPC1
931	P002	333333.3333	Strasbourg	3	Jean-Pierre	Sauvage	10/21/1944	0000-00-00	France	FR	Paris	NULL	NULL	NULL	male	NPC2
932	P002	333333.3333	Evanston, IL	3	Sir J. Fraser	Stoddart	5/24/1942	0000-00-00	United Kingdom	GB	Edinburgh	NULL	NULL	NULL	male	NPC2
933	P002	333333.3333	Groningen	3	Bernard L.	Feringa	5/18/1951	0000-00-00	the Netherlands	NL	Barger-Compascuum	NULL	NULL	NULL	male	NPC2
934	P002	500000	Cambridge, MA	2	Oliver	Hart	10/9/1948	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC1
935	P002	500000	Cambridge, MA	2	Bengt	Holmström	4/18/1949	0000-00-00	Finland	FI	Helsinki	NULL	NULL	NULL	male	NPC6
927	P003	1000000	Tokyo	1	Yoshinori	Ohsumi	0000-00-00	0000-00-00	Japan	JP	Fukuoka	NULL	NULL	NULL	male	NPC3
928	P003	500000	Seattle, WA	2	David J.	Thouless	9/21/1934	0000-00-00	United Kingdom	GB	Bearsden	NULL	NULL	NULL	male	NPC1
929	P003	250000	Princeton, NJ	4	F. Duncan M.	Haldane	9/14/1951	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC1
930	P003	250000	Providence, RI	4	J. Michael	Kosterlitz	6/22/1942	0000-00-00	United Kingdom	GB	Aberdeen	NULL	NULL	NULL	male	NPC1
931	P003	333333.3333	Strasbourg	3	Jean-Pierre	Sauvage	10/21/1944	0000-00-00	France	FR	Paris	NULL	NULL	NULL	male	NPC2
932	P003	333333.3333	Evanston, IL	3	Sir J. Fraser	Stoddart	5/24/1942	0000-00-00	United Kingdom	GB	Edinburgh	NULL	NULL	NULL	male	NPC2
933	P003	333333.3333	Groningen	3	Bernard L.	Feringa	5/18/1951	0000-00-00	the Netherlands	NL	Barger-Compascuum	NULL	NULL	NULL	male	NPC2
935	P003	500000	Cambridge, MA	2	Oliver	Hart	10/9/1948	0000-00-00	United Kingdom	GB	London	NULL	NULL	NULL	male	NPC6

Final Comments and Difficulties Faced

- Cleaning the Data
  - Foreign names
  - Countries that no longer existed
- ETL process
  - Stage 2
- Tableau and Data Visualization
  - Unfamiliar software
  - Applying effective data visualization

References

Kuzmenko, Maryna. (2016). *Nobel Prize - Dataset with Information about Prizes, Laureates and Countries*. Harvard Dataverse, V1, UNF:6:McdDh+ldUTGgZDs5XVOQUA== [fileUNF]. https://doi.org/10.7910/DVN/AGAFAQ

The Nobel Foundation. (2022, April 13). *The Nobel Prize*. NobelPrize.Org. https://www.nobelprize.org/