OLAP Queries and BI Dashboard

CSI 4142 - Fundamentals of Data Science

Winter 2022

School of Electrical Engineering and Computer Science

University of Ottawa

Course Coordinator: Dr. Herna L Viktor

Teaching Assistants: Anuhbav Chharbra, Nicolas Fleece, Paul Mvula

Group 25:

Lilian Ly, 8262186

Jonathan Brar, 8209351

Logan Rose, 300059034

Submission Date: April 8, 2022

Table of Contents

OLAP Queries

Standard OLAP Operations (9 queries)

Drill Down	

Get the average deaths drilled down by quarter, year, and decade	SELECT
	  M.name,
	  M.year,
	  AVG(F.avg_deaths)
	FROM
	  fact F,
	  event E,
	  month M
	WHERE F.month_key = M.month_key
	GROUP BY (M.name, M.year);

Roll Up	

Get the Average HDI rolled up for each country, region, continent.	SELECT
	  AVG(F.hdi),
	  C.short_name,
	  C.continent
	FROM
	  fact as F,
	  country as C,
	  event as E
	WHERE
	  F.country_key = C.country_key
	  AND F.event_key = E.event_key
	GROUP BY ROLLUP (C.continent, C.short_name);

Slice	

Get the average HDI in all 9 countries.	SELECT
	  AVG(F.hdi),
	  C.short_name,
	  M.name,
	  M.year
	FROM
	  fact as F,
	  country as C,
	  month as M
	WHERE
	  F.country_key = C.country_key
	  AND F.month_key = M.month_key
	GROUP BY (C.short_name, m.month_key)
	ORDER BY M.year, M.month_number;

Dice	

Compare the average number of deaths and average number of births that took place in Canada and the US after 2008	SELECT
	    avg(F.avg_births) as births,
	    avg(F.avg_deaths) as deaths,
	    C.short_name,
	    M.year
	FROM fact as F,
	  country as C,
	  Month as M
	WHERE
	  F.country_key = C.country_key AND
	  F.month_key = M.month_key
	  AND C.alpha_code in ('CAN', 'USA')
	  AND year > 2008
	GROUP BY (C.short_name, m.year);

Get the average HDI of lower middle and low income countries before 2010.	SELECT
	  AVG(F.hdi),
	  C.short_name,
	  M.name,
	  M.year
	FROM
	  fact as F,
	  country as C,
	  month as M
	WHERE
	  F.country_key = C.country_key
	  AND F.month_key = M.month_key
	  AND C.income_group in ('Lower middle income', 'Low income')
	  AND M.year<2010
	GROUP BY (C.short_name, m.month_key);

Combining OLAP Operations	

Compare the average number of deaths that took place between the years 2005 and 2015 in Asia, North and South America.	SELECT
	  AVG(F.avg_deaths),
	  C.short_name,
	  C.continent
	FROM
	  fact as F,
	  country as C,
	  Month as M
	WHERE
	  F.country_key = C.country_key
	  AND F.month_key = M.month_key
	  AND M.year > 2005
	  AND M.year < 2015
	  AND C.continent in ('AS','NA','SA')
	GROUP BY ROLLUP (C.continent, C.short_name);

Get the average number of births, rolled up monthly, quarterly, yearly, and per decade (Roll up) Looking only at instances where the maternal leave benefits ate less than 50% for a given year (Slice).	SELECT AVG(F.avg_births), M.name, M.quarter, M.year, M.decade FROM fact as F, month as M, quality_of_life as Q WHERE F.month_key = M.month_key AND F.month_key = M.month_key AND F.quality_of_life_key = Q.quality_of_life_key and Q.maternal_leave_benifits < 50 GROUP BY ROLLUP (M.decade, M.year, M.quarter, M.name);
Get the Average HDI of countries that spend more than 5% of their budget on Health expenditure, by continent.	SELECT DISTINCT C.short_name, C.continent, AVG(F.hdi) as HDI FROM fact as F, country as C, health as H, month as M WHERE F.country_key = C.country_key AND F.health_key = H.health_key AND F.month_key = M.month_key AND H.health_expenditure > 5 GROUP BY ROLLUP(C.continent, C.short_name) ORDER BY C.short_name;
Get the Average number of births and deaths per country rolled up monthly, quarterly, yearly, and per decade where the average UHC coverage index is between 85 and 86.	SELECT DISTINCT AVG(avg_births), AVG(avg_deaths), M.name, M.quarter, M.year, M.decade, C.short_name FROM fact as F, month as M, health as H, country as C WHERE F.month_key = M.month_key AND F.country_key = C.country_key

	AND AVG(H.uhc_service_coverage_index) > 75 GROUP BY C.short_name, ROLLUP (M.decade, M.quarter, M.year, M.name);

Explorative Operation (3 queries)

Iceberg	

Find the five years with the highest average births.	SELECT
	AVG(F.AVG_BIRTHS) AVG_BIRTHS,
	M.year
	from
	Fact F,
	Month M
	WHERE
	F.month_key = M.month_key
	GROUP BY
	(M.year)
	ORDER BY
	AVG_BIRTHS
	LIMIT
	5;

Windowing	
The HDI of a given country in a given year compared to the average HDI for the given country over the whole time set	SELECT C.short_name, M.year, F.hdi, Round((AVG(F.HDI) over (PARTITION BY C.short_name)), 2) as country_avg_hdi from Fact F, Country C, Month M WHERE F.country_key = C.country_key and F.month_key = M.month_key GROUP BY C.short_name, M.year, F.hdi ORDER BY M.year
Window Clause	
Get the Average Number of deaths and people affected based on disaster type.	SELECT distinct(E.event_type), ROUND(avg(E.total_deaths) over W, 2) as AVG_DEATHS, ROUND(avg(E.total_affected) over W, 2) as AVG_AFFECTED

	FROM
	  fact F,
	  event E
	WHERE F.event_key = E.event_key
	WINDOW W as (
	    PARTITION BY e.event_type
	    ORDER BY e.event_type
	);

BI Dashboard and Information Visualization

Drill down

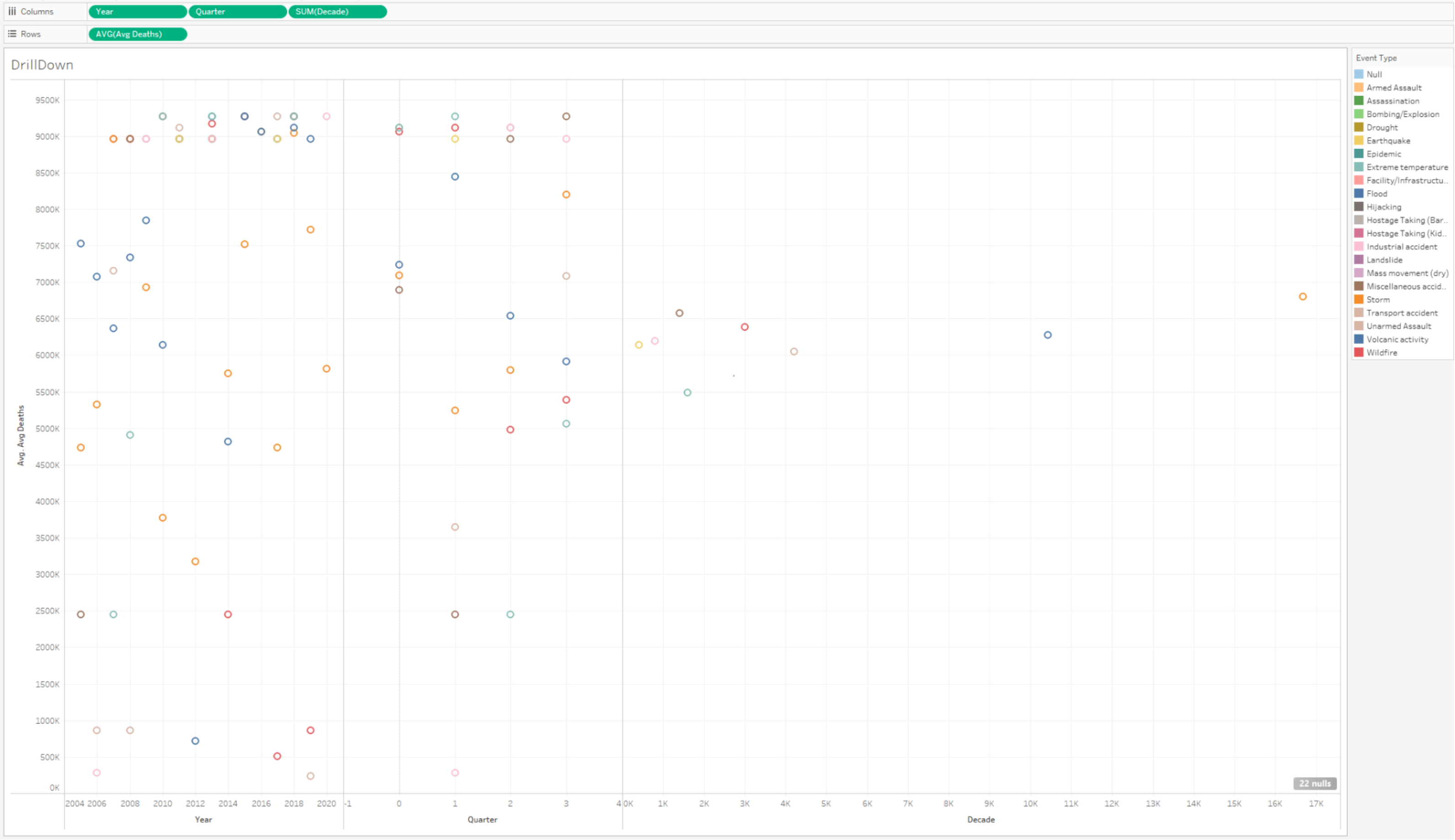
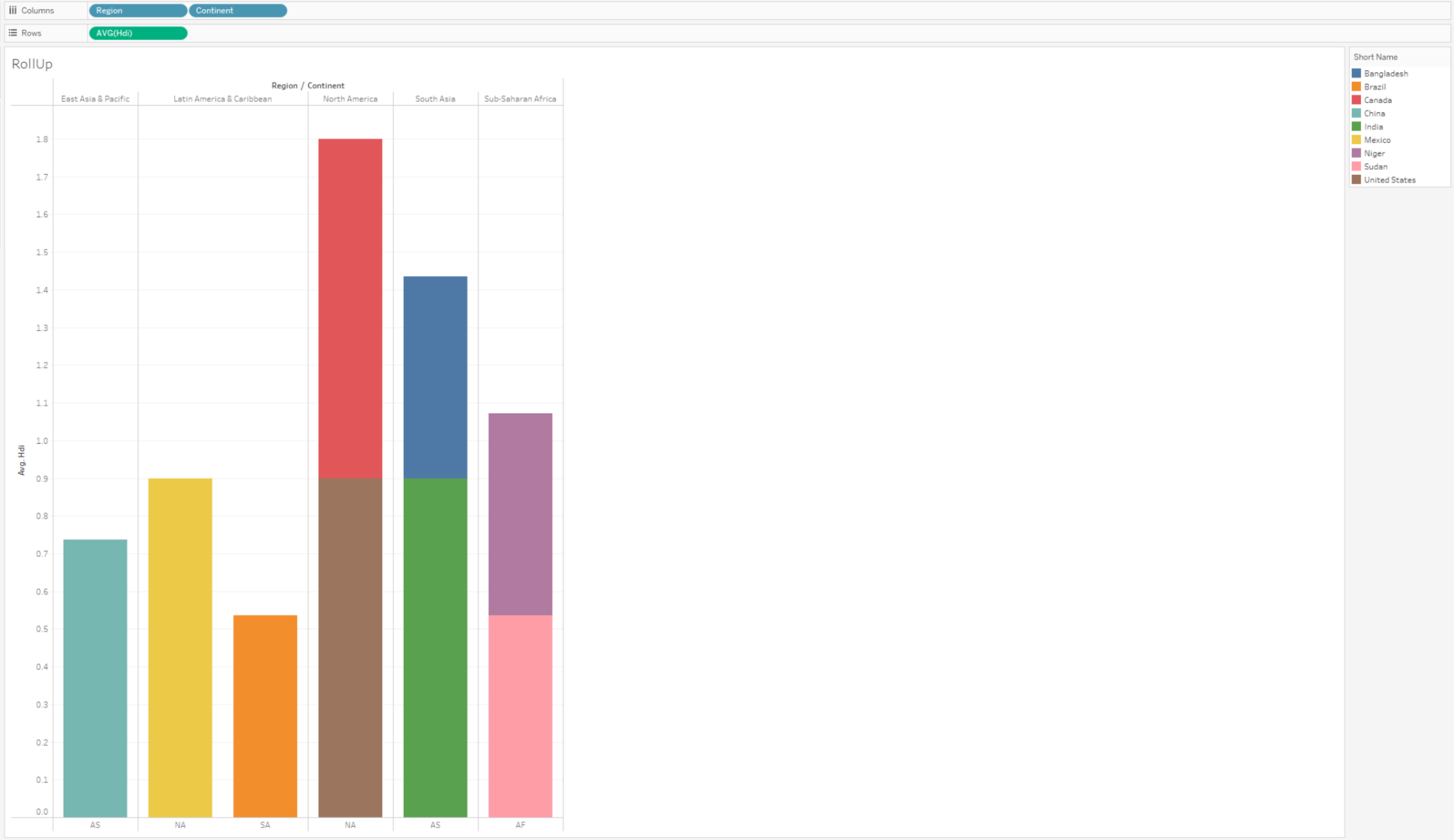


Figure 1. Get the average deaths drilled down by quarter, year, and decade

Roll up

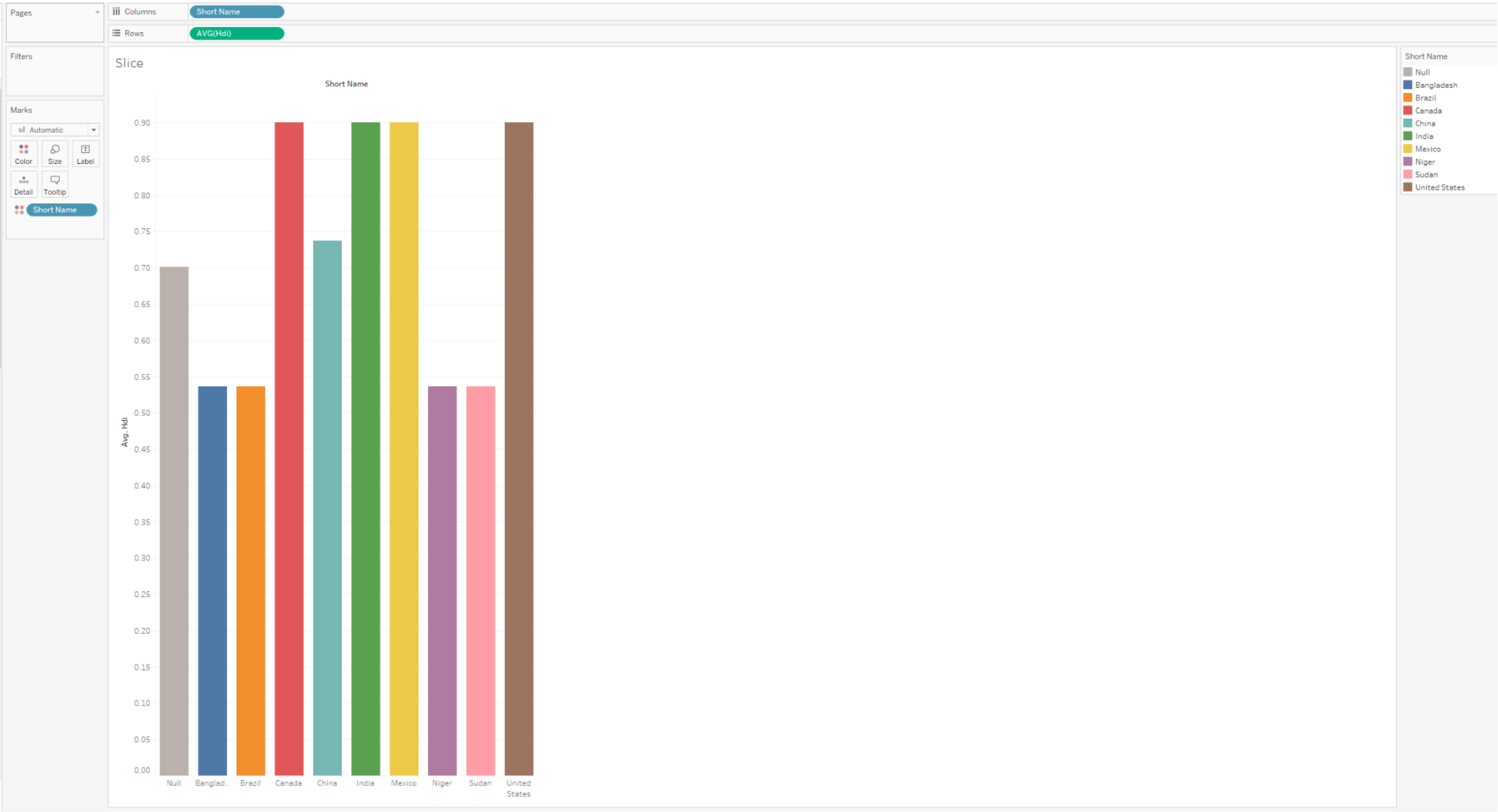


Columns: Region, Continent

Rows: AVG(Hdi)

Figure 2. Get the Average HDI rolled up for each country, region, continent.

Slice



Columns: Short Name

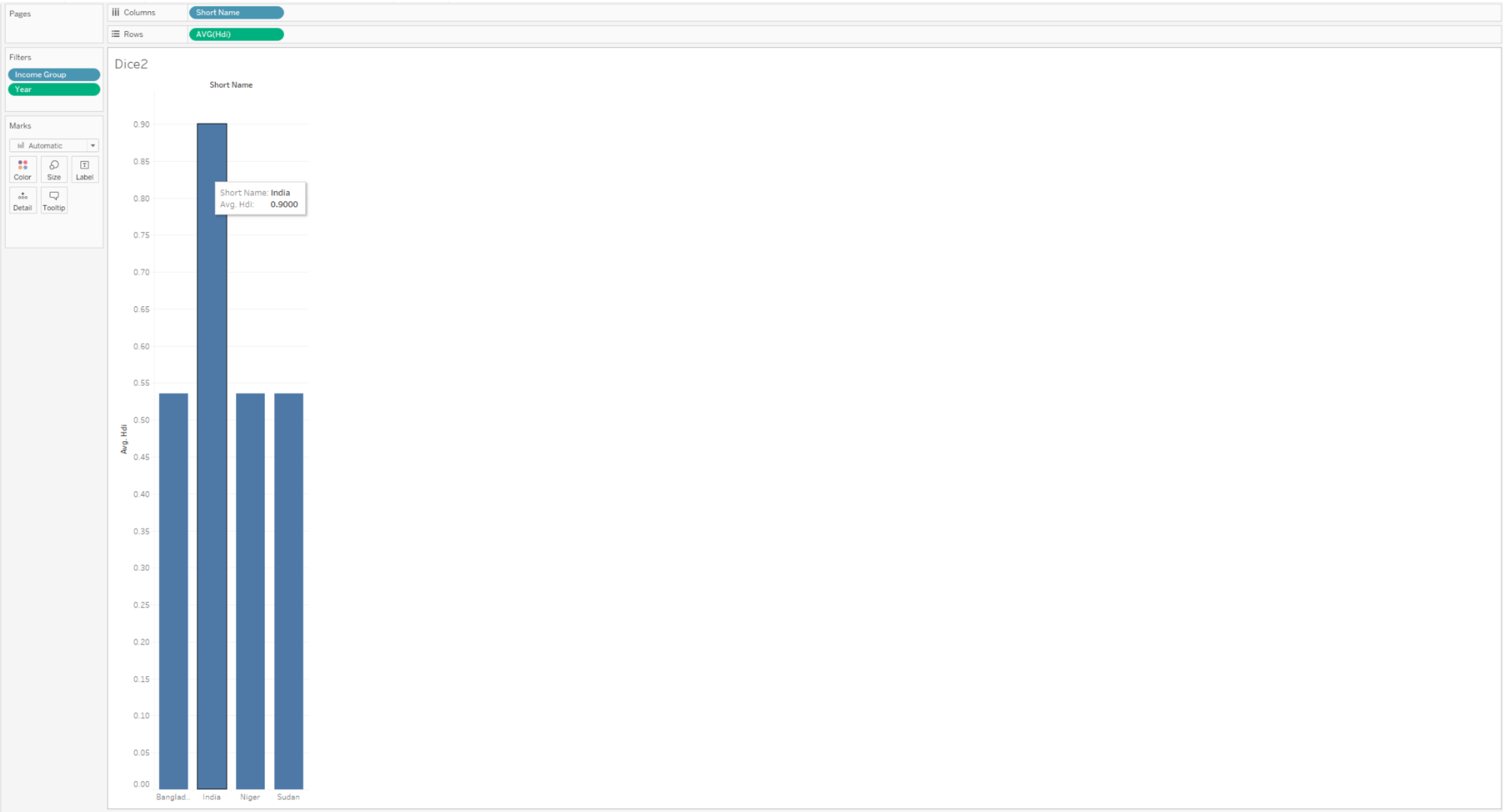
Rows: AVG(Hdi)

Marks: Automatic

Color: Short Name

Figure 3. Get the average HDI in all 9 countries.

Dice



Columns: Short Name

Rows: AVG(Hdi)

Filters: Income Group, Year

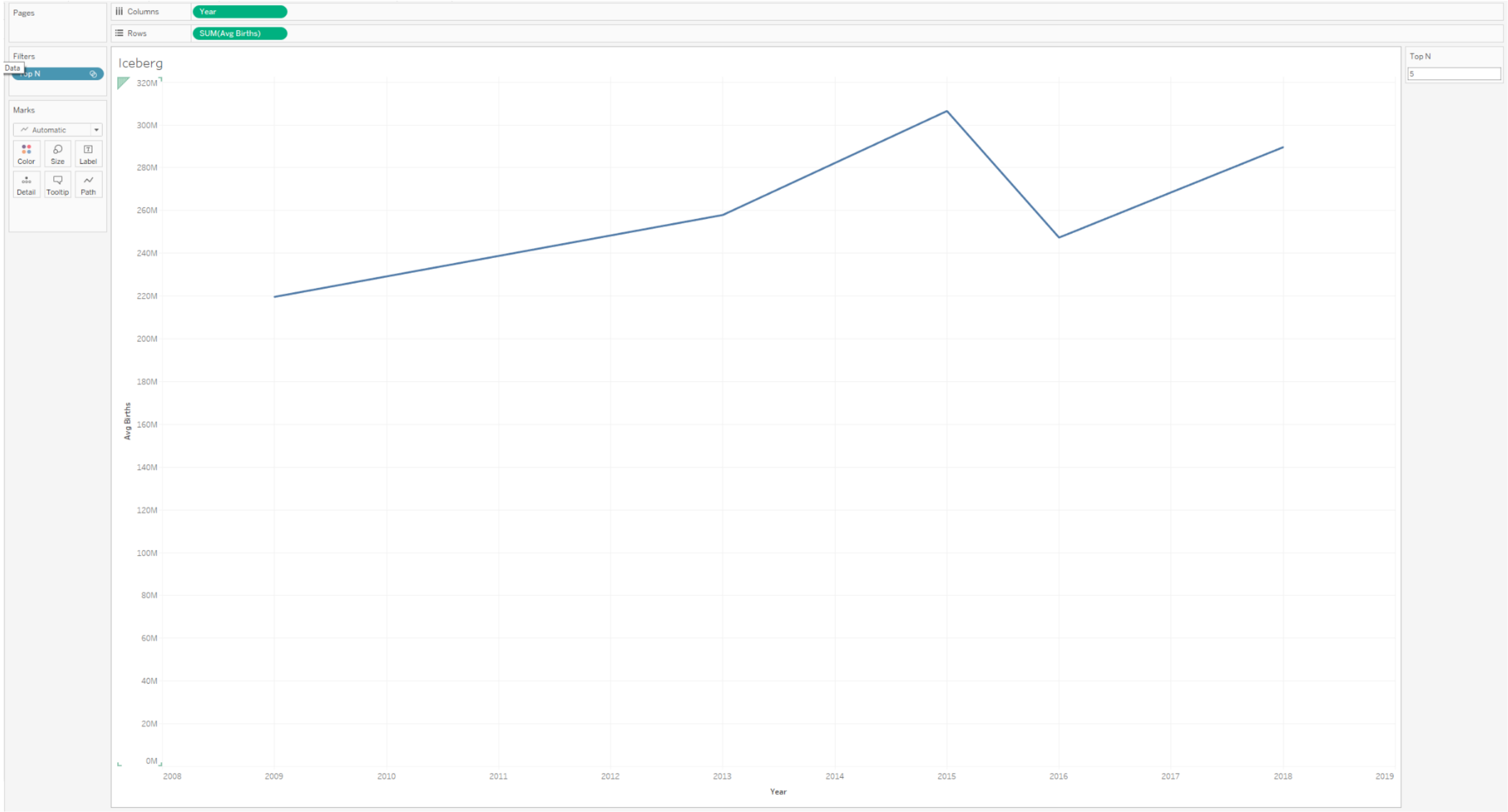
Marks: Automatic

Dice2

Short Name: India
Avg. Hdi: 0.9000

Figure 4. Get the average HDI of lower middle and low income countries before 2010.

Top N



Columns: Year

Rows: SUM(Avg Births)

Top N: 5

Figure 5. Find the five years with the highest average births.

Team Work Plan

Deliverable checklist	Responsible team member(s)	Expected completion date	Actual completion date	Estimated time (hours) to complete	Actual time (hours) to complete
OLAP queries					
Drill down, roll up	Logan	April 3, 2022	April 3, 2022	20 mins	30 mins
Slice	Logan, Lilian, Jonathan	April 3, 2022	April 4, 2022	15 mins	15 mins
Dice	Logan, Lilian Jonathan	April 3, 2022	April 4, 2022	20 mins	30 mins
Composite OLAP - 4 queries	Logan, Lilian, Jonathan	April 3, 2022	April 8, 2022	40 mins	50 mins
Icebergs	Logan	April 3, 2022	April 4, 2022	15 mins	15 mins
Windowing - partition	Logan	April 3, 2022	April 3, 2022	15 mins	20 mins
Window	Logan	April 3, 2022	April 1, 2022	15 mins	15 mins
BI dashboard					
Design of data mart	Jonathan	March 28, 2022	March 28, 2022	30 mins	30 mins
Importing data	Jonathan	March 29, 2022	March 29, 2022	15 mins	10 mins
OLAP queries	Jonathan	April 4, 2022	April 4, 2022	30 mins	30 mins
Figures	Jonathan	April 4, 2022	April 4, 2022	45 mins	1h
Other tasks - please specify					
Creating Events script	Lilian	March 28, 2022	March 31, 2022	1h	1h30
Report	Lilian	April 2, 2022	April 5, 2022	1h30	2h
Submission	Lilian	April 2, 2022	April 5, 2022	15 min	15 min

Github Repository

OLAP Script File:
https://github.com/Logan-Rose/Data-Science-Project/blob/main/data-mart/olap_queries.sql

Repository Link: https://github.com/Logan-Rose/Data-Science-Project