



OBJECTIVE

This project is used to extract information from agriculture data and to give suggestions regarding crops and make future decisions so that agriculture can be carried out in a planned manner.







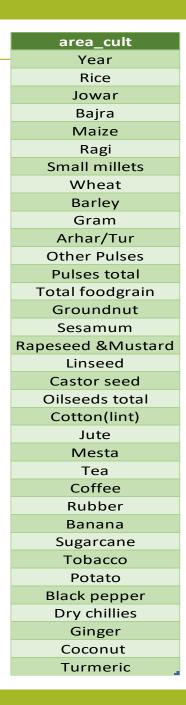
TECHNOLOGIES USED

- ❖ AWS S₃
- * AWS Glue
- * AWS Crawler
- AWS Wrangler
- AWS Redshift
- * AWS Lambda
- PySpark
- Spark SQL

SCHEMA DIAGRAM

STATES
YEAR
Crop
Area
Production
Annual_Rain
Avg_Temperature
Price
Cost_per_Hectare
Cost_per_quintal
Suicides

Export





PROBLEM STATEMENTS



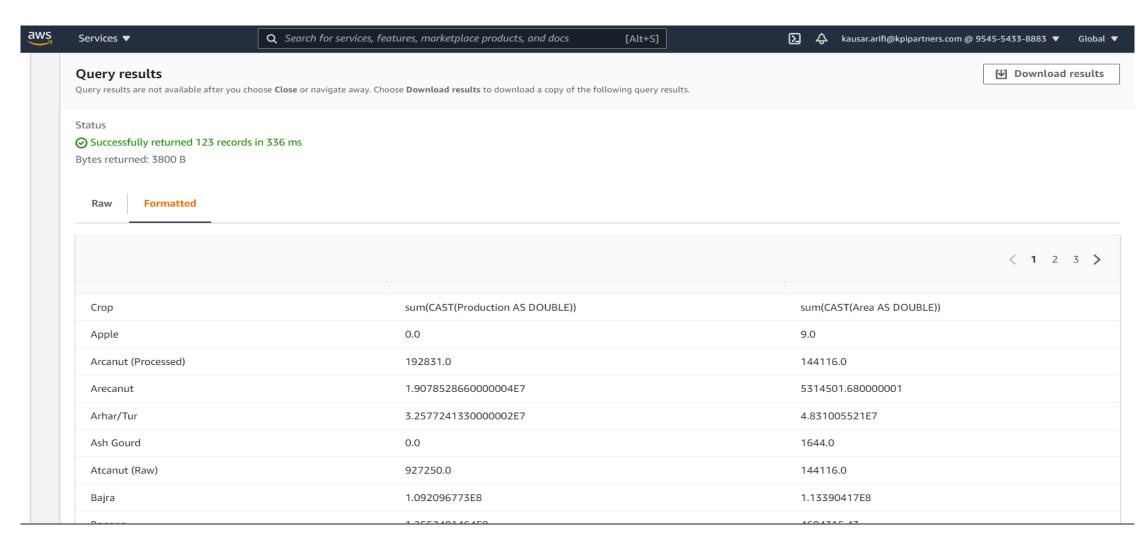
- 1. Finding trends in crops in terms of production, area.
- 2. State wise suicide rate of the farmers
- 3. Which year has Highest cultivation of Pulses
- 4. Which year has Highest cultivation among rice, wheat and coffee.
- 5. Which state yields highest crop per year?
- 6. Finding how different factors like temperature affect production
- 7. Year wise suicide rates of farmers
- 8. Which state receives highest rainfall in which year and what is the annual rainfall?
- 9. Crop wise cost of cultivation versus cost of production
- 10. Highest crop production in particular season.

1. Finding trends in crops in terms of production, area.



aws Ser	rvices ▼	Q S	earch for serv	vices, feature	es, marketpl	ace products, ai	nd docs	[Alt+S		\$\triangle kausar.arifi@kpipartners.com @ 9	545-5433-8883 ▼ Oregon ▼ Suppo	ort 🔻
Agricul	lture_K1	ß								Job has not been saved	Save Delete Run	
Visual	Script	Jol	o details	Runs	Schedules	5						
Source •	☆ Transfor	m 🔻	齿 Target ▼	⇔ Undo	♂ Redo	T Remove	•	Q	20	Node properties Transfor	m Output schema	3
			Data source - S3 bu S3 bucket	ucket 🕢						Data preview		
										Associate an alias with each inpu Enter valid SQL table names to associa		
			Transform - SQL Co	nde						Input sources	Spark SQL aliases	
		- SQL	Spark SQL	⊗						S3 bucket	agri	
										Code block Enter SQL code to add to your job.		
		0	Data target - S3 bu Amazon S3	ocket 🕢						1 SELECT Crop, sum(Area), 2 FROM agri 3 GROUP BY Crop 4 ORDER BY Crop	sum(Production)	-









aws Sei	ervices v Q	Search for se	rvices, featur	es, marketplo	ace products, a	nd docs	[Alt+S	1 2	⇔ kausar.a	nrifi@kpipartners.com @	9545-5433-8883	S ▼ Oregon	▼ Support ▼
Agricu	lture_KQ2					Last Sa	ved at 10	/20/2021	, 1:23:09 PM	Save	l session	Delete	Run
Visual	Script Jo	ob details	Runs	Schedules	S								
□ Source ↓	☆ Transform	≛ Target	⊅ Undo	♂ Redo	T Remove	Ф	Q	Œ	Node prop	perties Transfo	rm Outp	out schema	×
									Data previ	iew			
			Data source - S3 bucket	S3 bucket						an alias with each inpo QL table names to assoc		cFrames.	^
									Input source	ces	Spark SQL	_ aliases	- 1
			Transform - Si Spark SQL	QL Code					S3 bucke	t	Agri		
			Data target -: Amazon S3	33 bucket					1 SELECT 2 FROM A 3 GROUP	ode to add to your job. T STATES, sum(Suic	ides) AS Tot	al_Suicides,	count(DIS

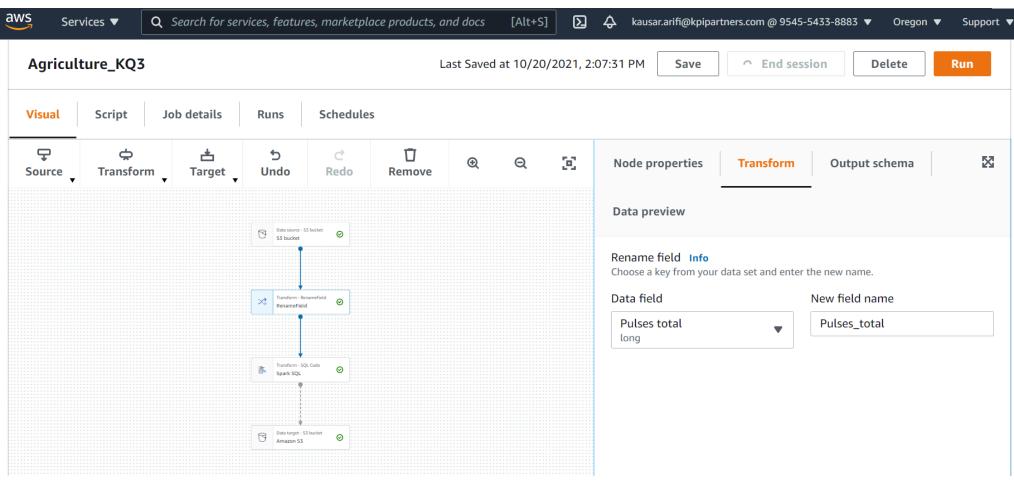




Services ▼	Q Search for services, features, marketplace p	roducts, and docs [Alt+S]	Σ		345-5433-8883 ▼	Global ▼
Query results Query results are not available after you choos	se Close or navigate away. Choose Download results to download	d a copy of the following query results.			Download r	results
Status						
Successfully returned 34 records in 3	39 ms					
Bytes returned: 1076 B						
Raw Formatted						
					<	1 >
STATES	Total	Suicides	Total_Years	Suicide_Rate		
ANDAMAN & NICOBAR ISLANDS	0.0		8	0.0		
ANDHRA PRADESH	4253	28.0	15	28355.2		
ARUNACHAL PRADESH	790.0		15	52.66666666666664		
ASSAM	0.0		15	0.0		
BIHAR	10929	9.0	15	728.6		
CHANDIGARH	0.0		11	0.0		
CHHATTISGARH	6651:	3.0	15	4434.2		
DADRA AND NAGAR HAVELI	0.0		15	0.0		
GOA	117.0		11	10.636363636363637		

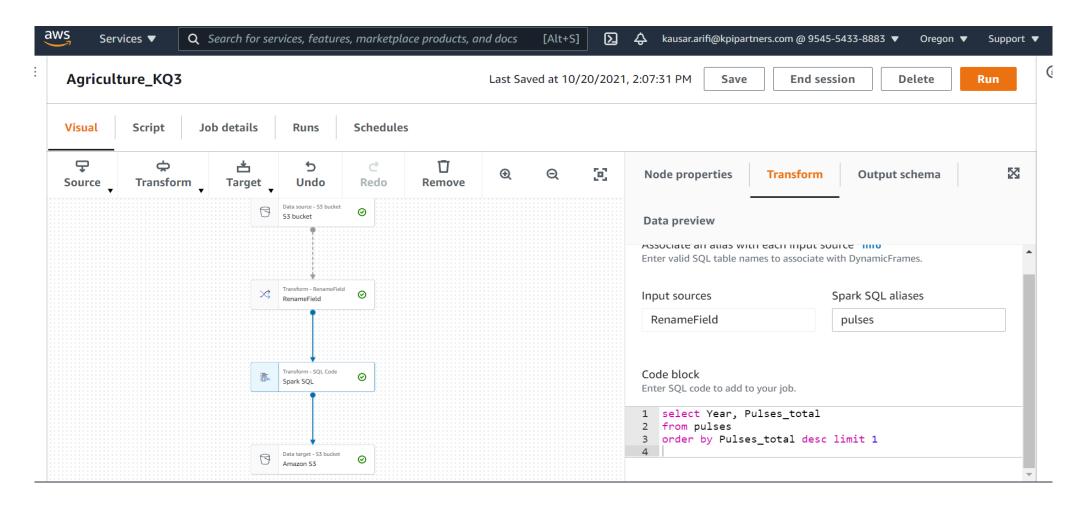
3. Which year has Highest cultivation of Pulses



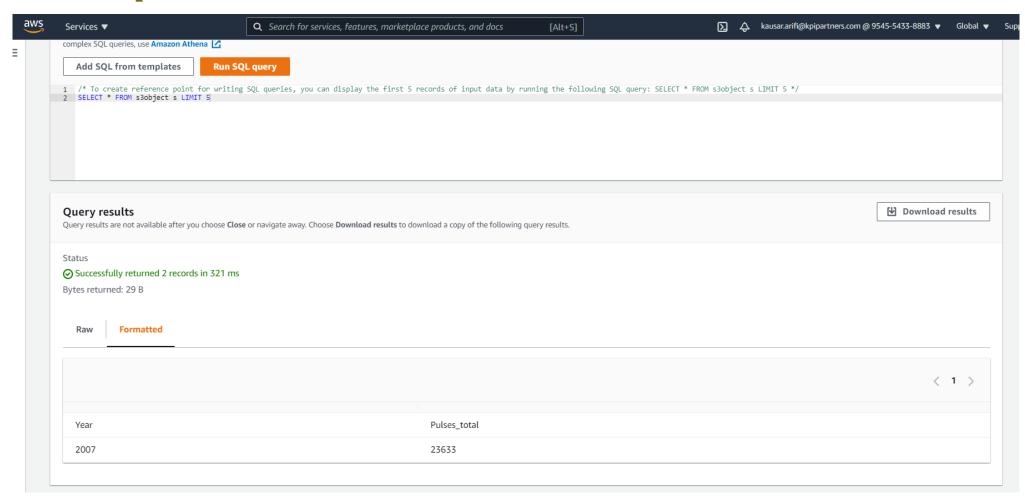




Cont...

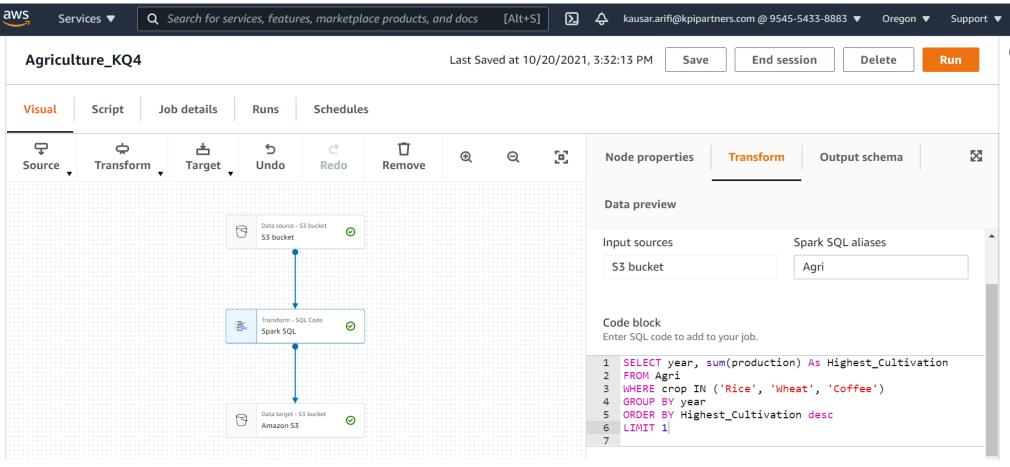






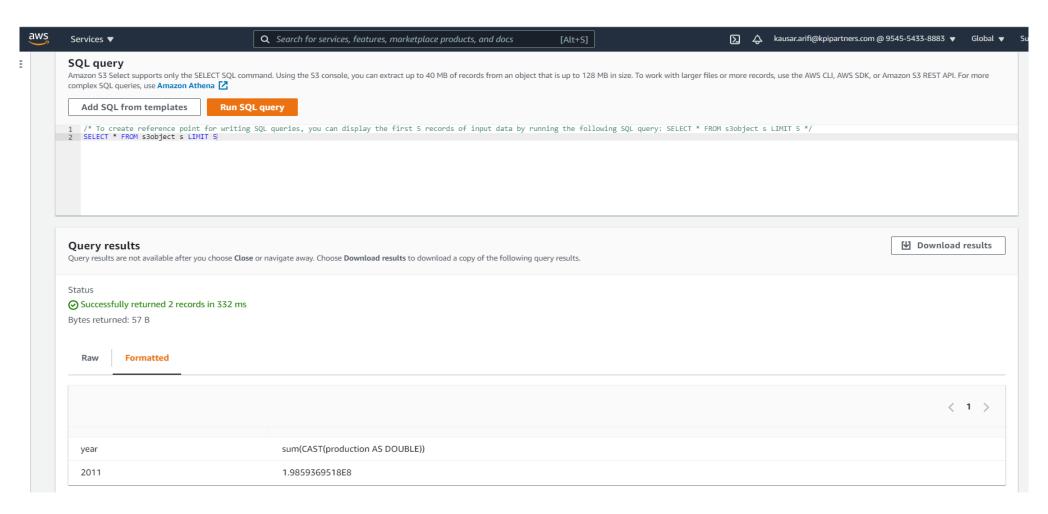
















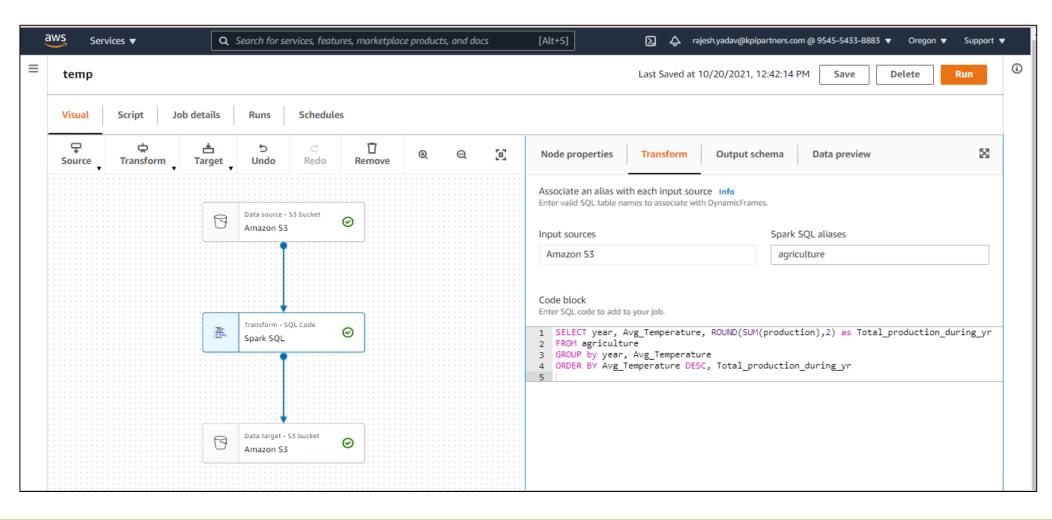
ō	Serv	vices ▼	Q	Search for s	ervices, featu	res, marketpla	ce produc	ts, and do)CS	[Alt+S]	∑ ⇔ raj	esh.yadav@kpipartners	com @ 9545-5433-8883 ▼	Oregon ▼ Support	•
=	high_yr	r									Last Saved at	10/20/2021, 7:57:3	7 PM Save De	elete Run	(1)
	Visual	Script	Job details	Runs	Schedule	s									
	Ç Source ▼	☆ Transform	± Target ↓	5 Undo	♂ Redo	Remove	•	Q	90	Node properties	Transform	Output schema	Data preview	×	
					e - Data Catalo e Data Catalo	(2)				Associate an alias wit Enter valid SQL table nar Input sources		h DynamicFrames.	k SQL aliases		
										AWS Glue Data Cata Code block Enter SQL code to add to		ag	riculture		
			Ē	Transform Spark SC	- SQL Code QL	⊗				3 4		re	ction), YEAR ORDER BY max(pro	duction) desc) as r	a
			9	Data targe Amazon	t - S3 bucket S3	0									



			< 1 >
NA24	STATES	max(production)	rankk
year			
2000	KERALA	5.536E9	1
2001	KERALA	5.479E9	1
2002	KERALA	5.709E9	1
2003	KERALA	5.876E9	1
2004	KERALA	6.001E9	1
2005	KERALA	6.326E9	1
2006	KERALA	6.054E9	1
2007	KERALA	5.641E9	1
2008	KERALA	5.802E9	1
2009	KERALA	5.667E9	1
2010	KERALA	5.286561E9	1
2011	TAMIL NADU	6.2009E9	1
2012	KERALA	5.799E9	1
2013	KERALA	5.921E9	1
2014	KERALA	5.947E9	1

6. Finding how different factors like temperature affect production







year	Avg_Temperature	Total_production_during_yr
2010	25.13	6.30760852537E9
2009	25.11	7.66049402481E9
2006	25.06	8.68191312378E9
2002	25.0	7.69695516264E9
2007	24.77	6.87944233858E9
2004	24.74	8.18946169486E9
2001	24.73	7.46554084027E9
2003	24.72	7.91797350473E9
2012	24.69	8.17105474603E9
2011	24.66	1.430890408713E10
2008	24.61	7.71701840216E9
2000	24.6	7.4497091271E9
2005	24.58	8.04375732974E9
2015	0.0	6935064.7
2014	0.0	8.66454063111E9
2013	0.0	1.290358863288E10



7. Year wise suicide rates of farmers

aws Se	ervices ▼	Q Searc	h for serv	ices, feature	s, marketpla	ace products, a	nd docs	[Alt+S]	Σ	⇔ kausar.arifi@kpipartner	rs.com @ 9545-5	433-8883 ▼	Oregon ▼	Support ▼
Agricu	ılture_KQ7	•							Last S	aved at 10/20/2021, 4:14:	:07 PM Sa	ave De	elete	Run
Visual	Script	Job de	tails	Runs	Schedules	5								
Source	☆ Transfor		åarget ↓	⇔ Undo	Ċ Redo	T Remove	•	Q	æ	Node properties	Transform	Output so	chema	×
										Data preview				
			5	Data source S3 bucket	- S3 bucket)				Associate an alias with e Enter valid SQL table names			nes.	Î
										Input sources	S _I	park SQL alia	ses	- 1
				Transform -	SQL Code					S3 bucket		Agri		
				spank sqc	- S3 bucket					Code block Enter SQL code to add to your selection of the	(Suicides) a	s Total_yea FROM Agri GROUP BY YE ORDER BY YE	AR	des



Services ▼	Q. Search for services, features, marketplace products, and docs [Alt+S]	
Query results Query results are not available after you choose Close or navigate away. Choose Download	d results to download a copy of the following query results.	
Status ② Successfully returned 17 records in 359 ms Bytes returned: 196 B		
Raw Formatted		
		⟨ 1 ⟩
YEAR	Total_yearly_suicides	
2015	0.0	
2014	138294.0	
2013	341826.0	
2012	437710.0	
2011	435725.0	
2010	497780.0	
2009	0.0	
2008	0.0	
2007	0.0	
2006	0.0	
2005	0.0	
2004	0.0	
2003	0.0	
2002	0.0	
2001	0.0	
2000	0.0	

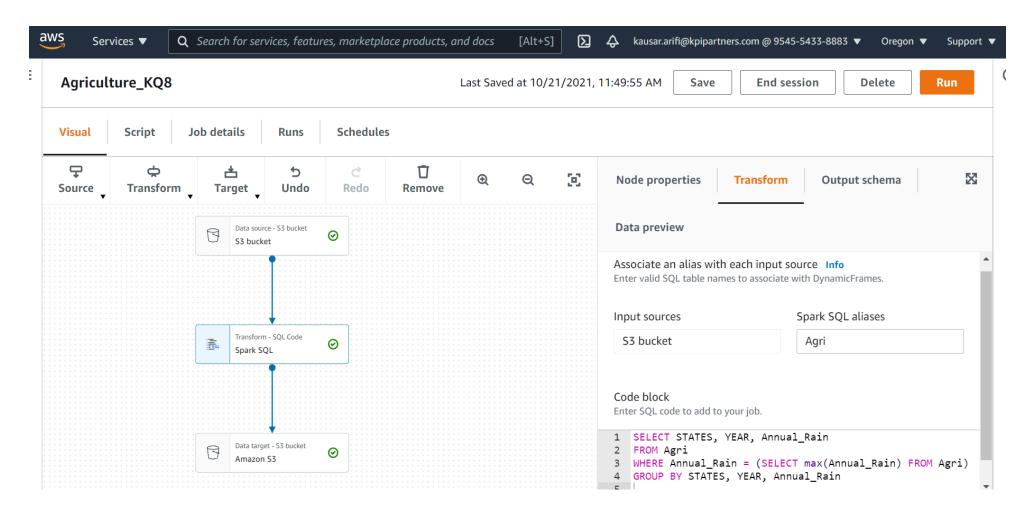
8. Which state receives highest rainfall in which year and what is the annual rainfall?



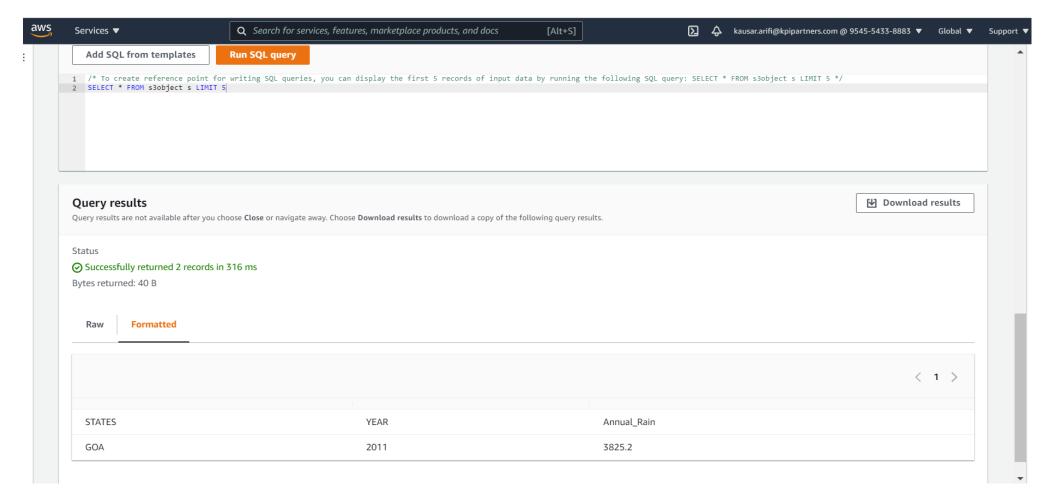
WS Ser	vices ▼ (Q Search for se	rvices, featur	es, marketplo	ace products, a	nd docs	[Alt+S] [2]	💠 kausar.arifi@kpipartn	ners.com @ 9545-543	33-8883 ▼ Orego	on ▼ Suppor
Agricul	ture_KQ8							Last S	Saved at 10/21/2021, 1:0	0:15 PM Sav	e Delete	Run
Visual	Script	Job details	Runs	Schedules	S							
□ Source	☆ Transform	± Target ↓	5 Undo	♂ Redo	Remove	•	Q	Œ	Node properties	Data source pro	operties - S3	X
			Data source -	S3 bucket					Output schema	Data preview		
			S3 bucket	⊘					S3 source type Info			
									Data Catalog tableS3 locationChoose a file or folde			
			Transform - So Spark SQL	QL Code					Database Choose a database.			
									agriculture_k		•	C
			Data target - S	53 bucket					Table			
			Amazon S3	⊗					data_set_csv		•	C

Cont...



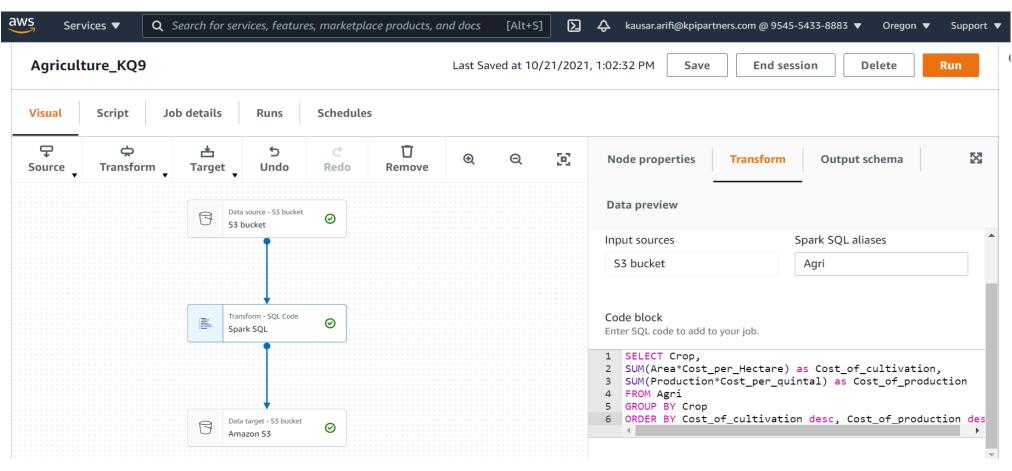




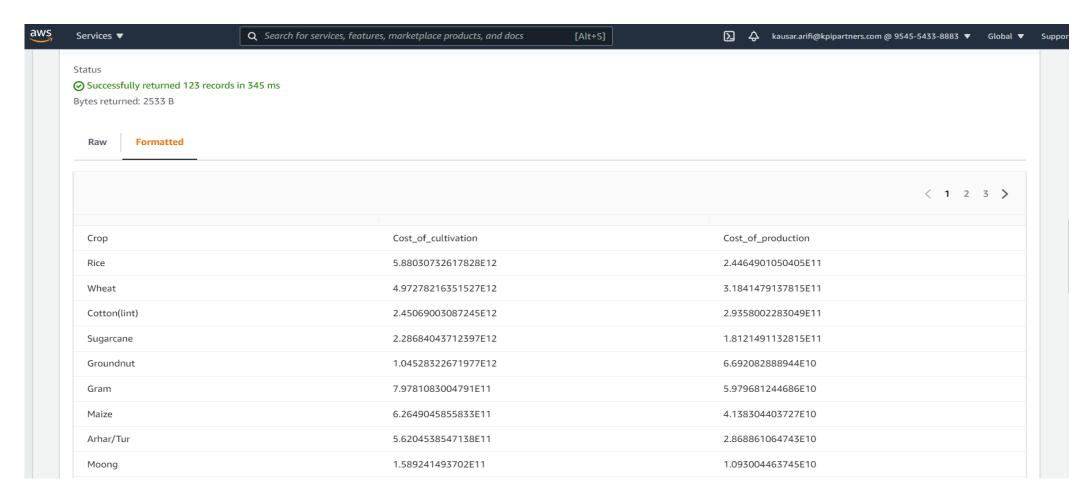






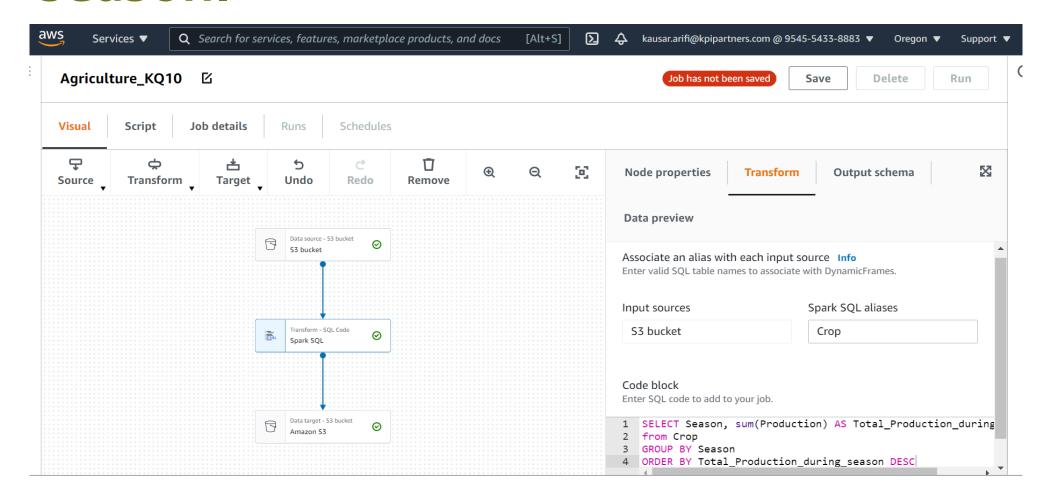






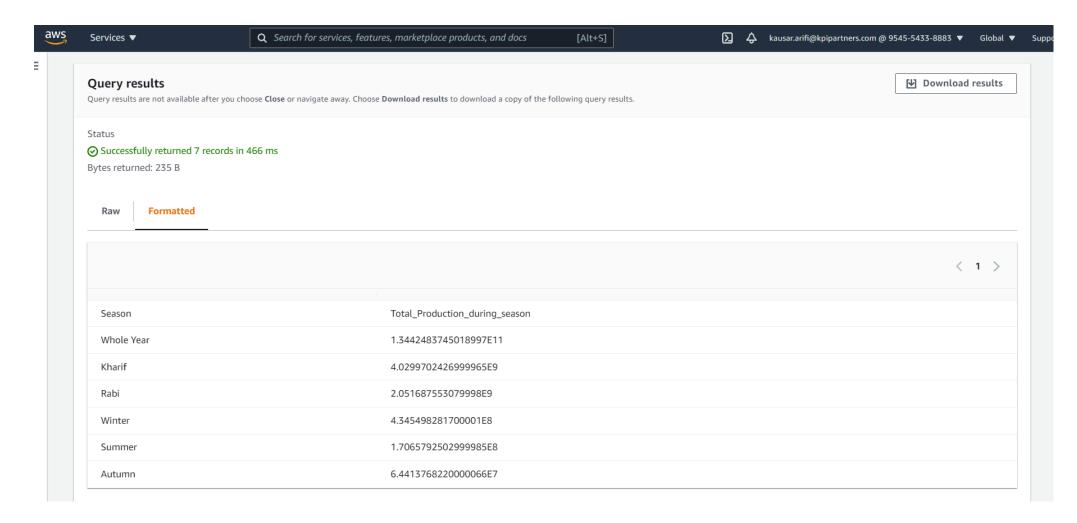
10. Highest crop production in particu season.





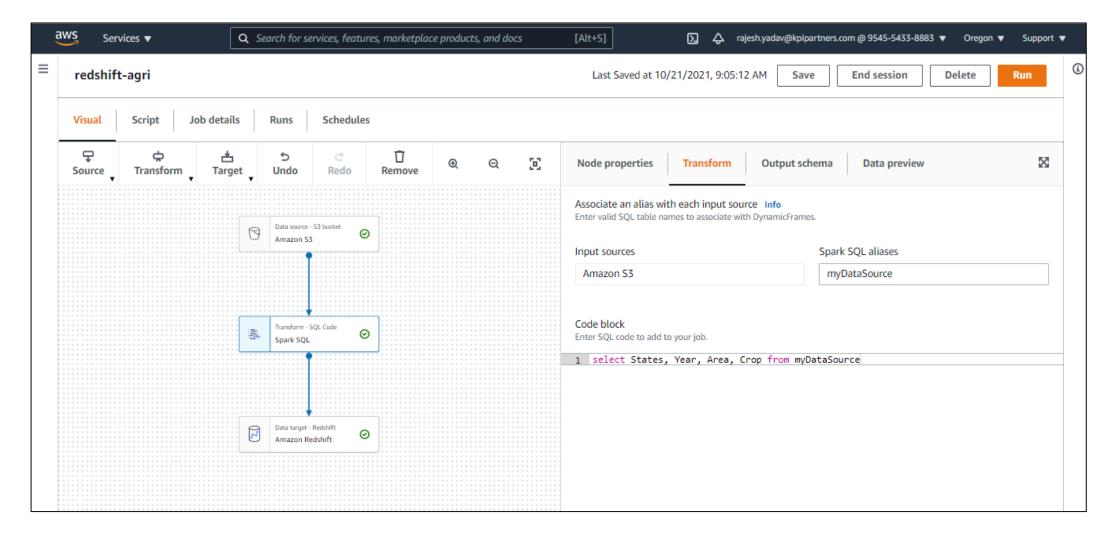






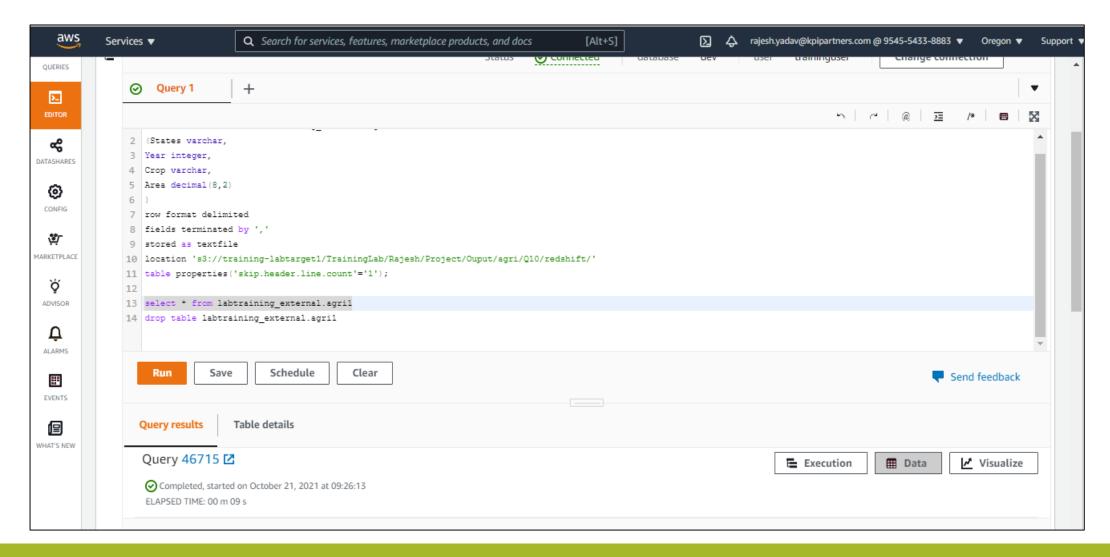


Visualization in Redshift -





Cont...

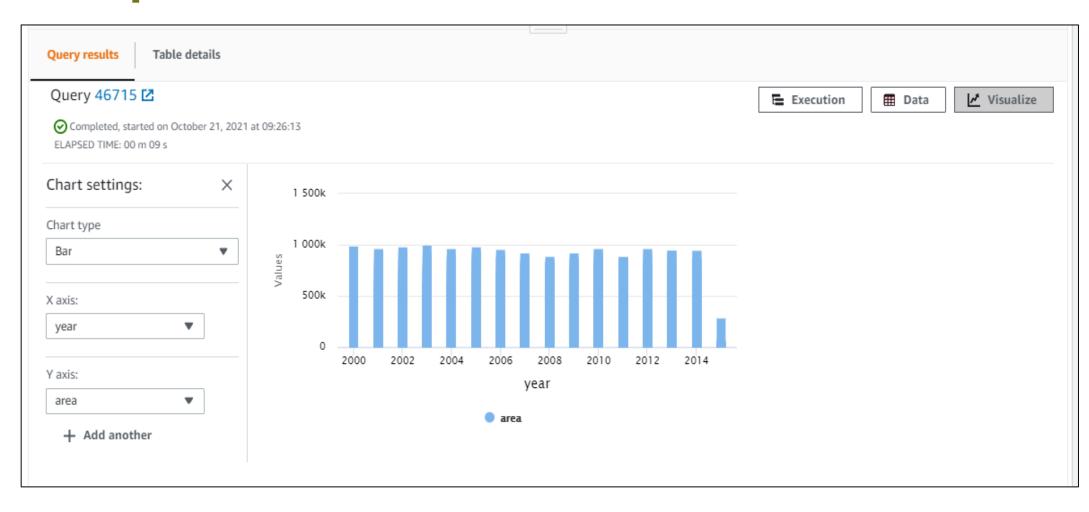




Cont..

Rows returned (11314)			Export ▼
Q Search rows		< 1	2 3 4 5 6 7 1132 >
states	▽ year	▽ crop	▼ area ▼
"ANDAMAN & NICOBAR ISLANDS"	2000	Arecanut	4354.00
"ANDAMAN & NICOBAR ISLANDS"	2000	"Other Kharif pulses"	670.00
"ANDAMAN & NICOBAR ISLANDS"	2000	Rice	10881.00
"ANDAMAN & NICOBAR ISLANDS"	2000	Banana	1707.00
"ANDAMAN & NICOBAR ISLANDS"	2000	Cashewnut	800.00
"ANDAMAN & NICOBAR ISLANDS"	2000	"Coconut "	25160.00
"ANDAMAN & NICOBAR ISLANDS"	2000	"Dry ginger"	388.00
"ANDAMAN & NICOBAR ISLANDS"	2000	Sugarcane	142.00
"ANDAMAN & NICOBAR ISLANDS"	2000	"Sweet potato"	83.00
"ANDAMAN & NICOBAR ISLANDS"	2000	Tapioca	200.00



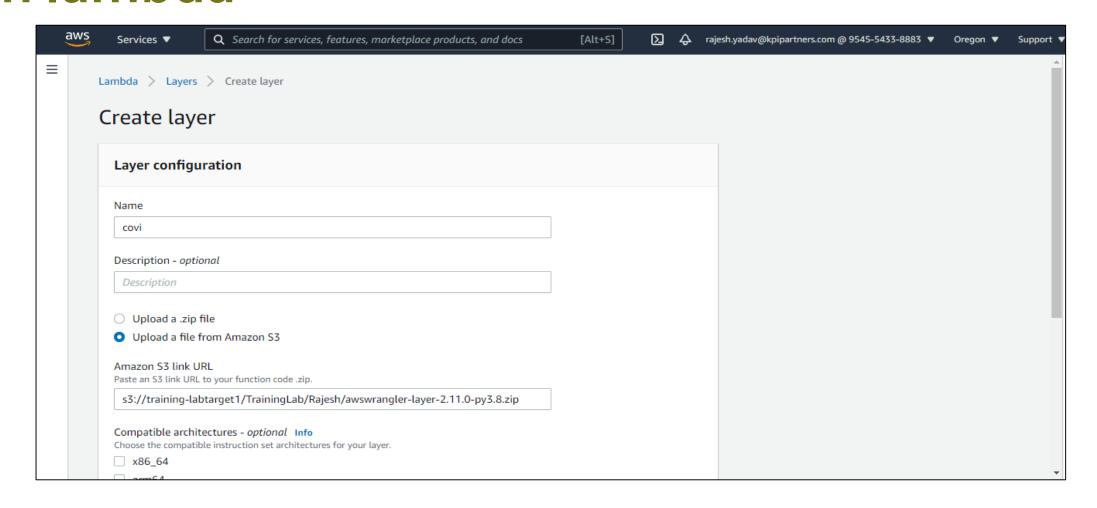




Using lambda function for transforming dataset and storing the modified dataset into S3 bucket

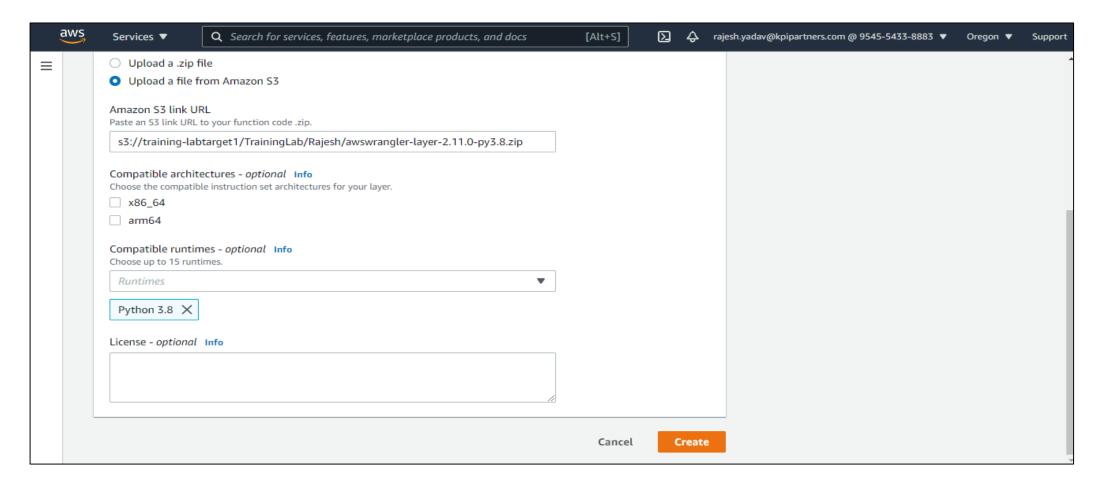
Creating layer for using AWS wrangler in lambda







Cont..



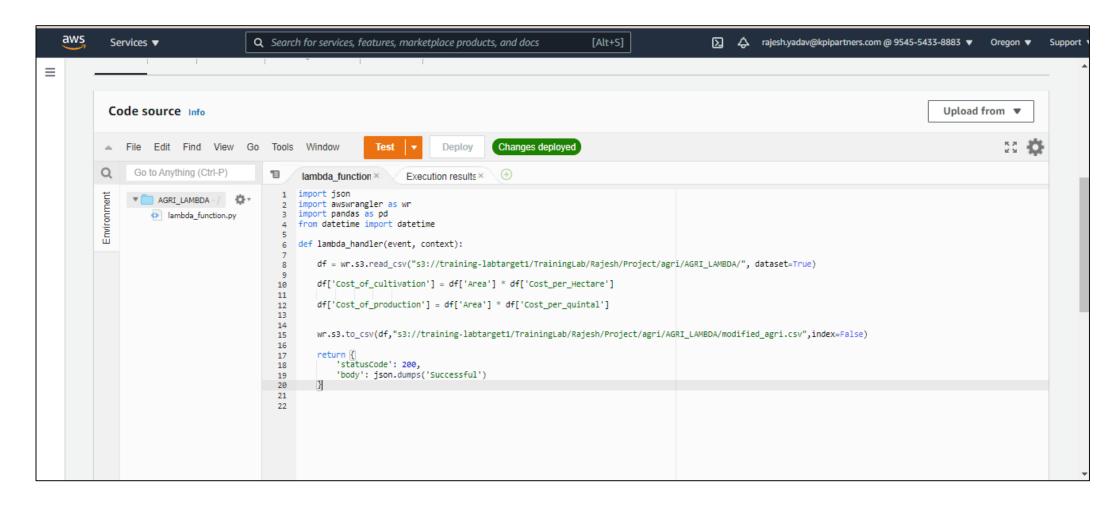


Cont..

aws	Services ▼	Q Search for services, features, marketplace products, and docs	[Alt+S]	№ 4	ra	jesh.yadav@kpipartners.com @ 9545-5433-8883 ▼	Oregon ▼	Support
=	O Upload a .zip f	file						
	Upload a file f	rom Amazon S3						
	Amazon S3 link U Paste an S3 link URL	RL to your function code .zip.						
	s3://training-lab	otarget1/TrainingLab/Rajesh/awswrangler-layer-2.11.0-py3.8.zip						
		ectures - optional Info ble instruction set architectures for your layer.						
	arm64							
	Compatible runtin	nes - <i>optional</i> Info times.						
	Runtimes		▼					
	Python 3.8 🗙							
	License - optional	Info						
			Canco	el Cre a	ite			

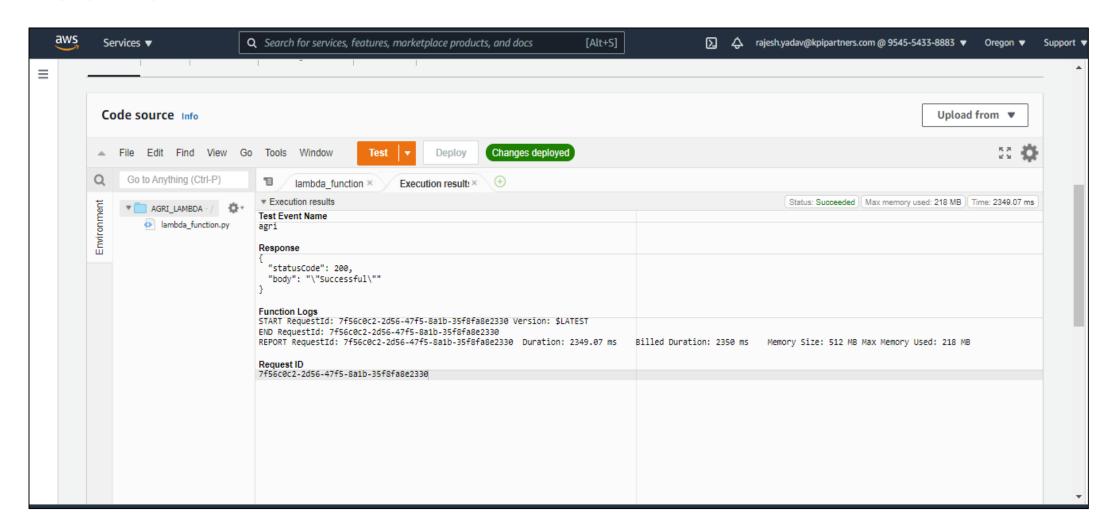
Transformation in lambda- Creating two additional columns







Cont...

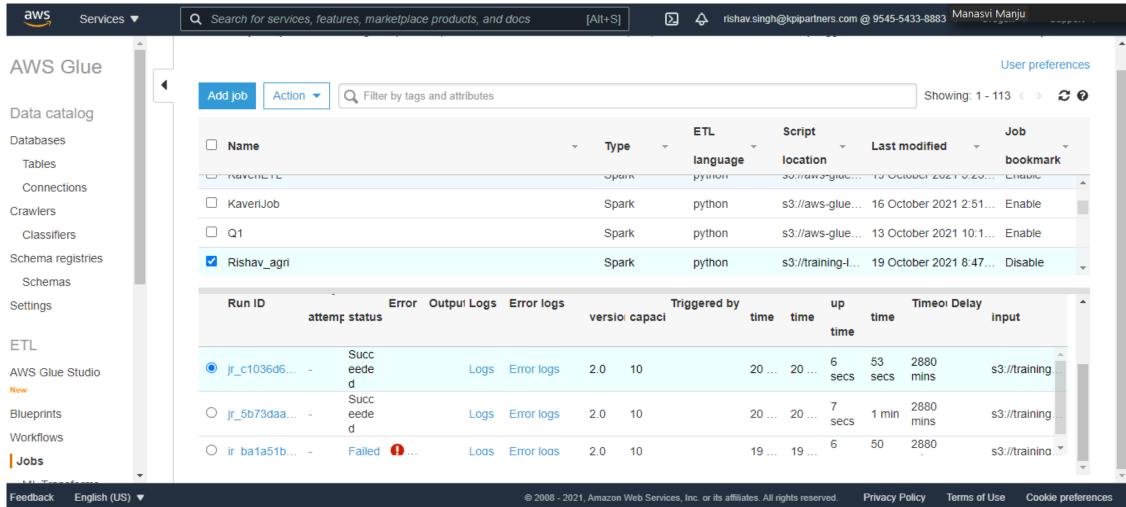




												< 1 2 3	4 5 6 7 20 >
STATES	YEAR	Crop	Area	Production	Annual_Rain	Avg_Temperature	Price	Cost_per_Hectare	Cost_per_quintal	Suicides	Export	Cost_of_cultivation	Cost_of_production
ANDAMAN & NICOBAR ISLANDS	2000	Arecanut	4354.0	7200.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Other Kharif pulses	670.0	449.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Rice	10881.0	32184.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Banana	1707.0	12714.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Cashewnut	800.0	219.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Coconut	25160.0	89000000.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Dry ginger	388.0	1220.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Sugarcane	142.0	3738.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Sweet potato	83.0	858.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2000	Tapioca	200.0	2277.0	2763.2	24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Arecanut	4354.0	7300.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Other Kharif pulses	420.0	200.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Rice	9801.0	27333.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Cashewnut	800.0	225.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Coconut	25205.0	89680000.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Dry ginger	390.0	1224.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Sugarcane	82.0	2380.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2001	Sweet potato	111.0	38.0	3080.9	24.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2002	Rice	10885.0	32111.66	2620.2	25.0	1130.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2002	Arecanut	4363.0	7350.0	2620.2	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2002	Banana	1737.0	12160.0	2620.2	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2002	Black pepper	550.0	116.0	2620.2	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ANDAMAN & NICOBAR ISLANDS	2002	Cashewnut	800.0	232.0	2620.2	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

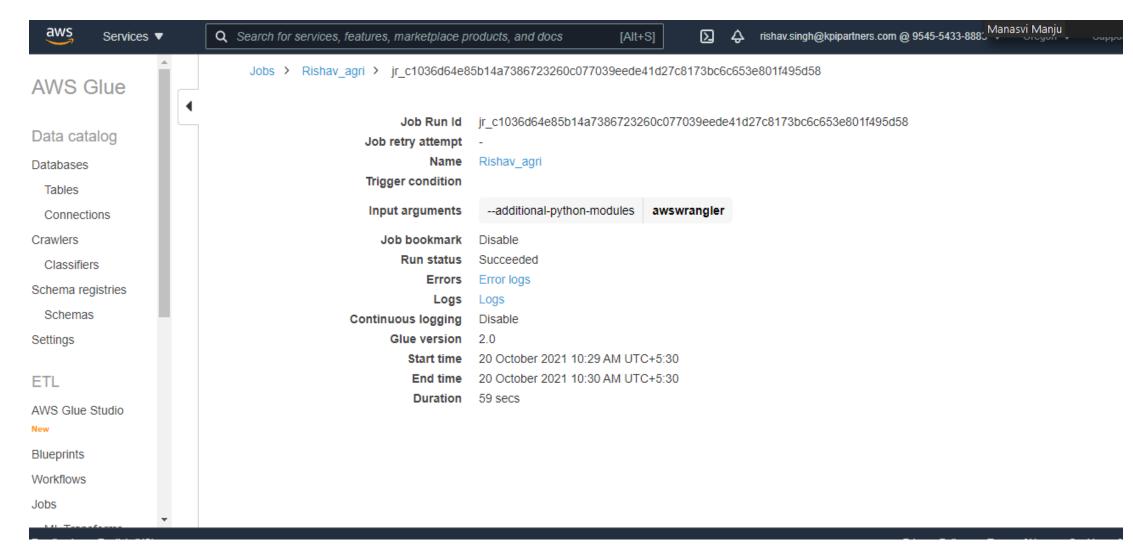
Adding Job





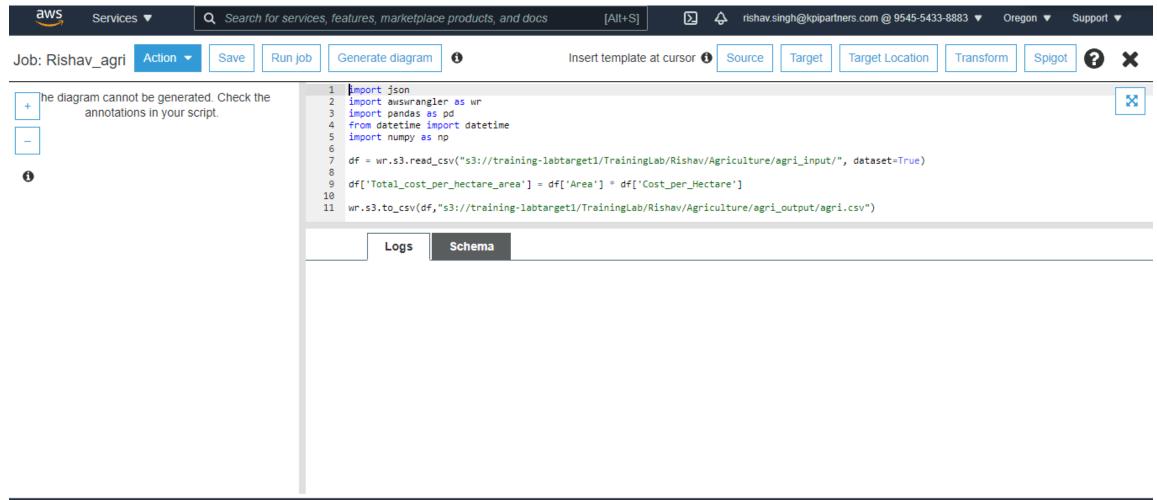




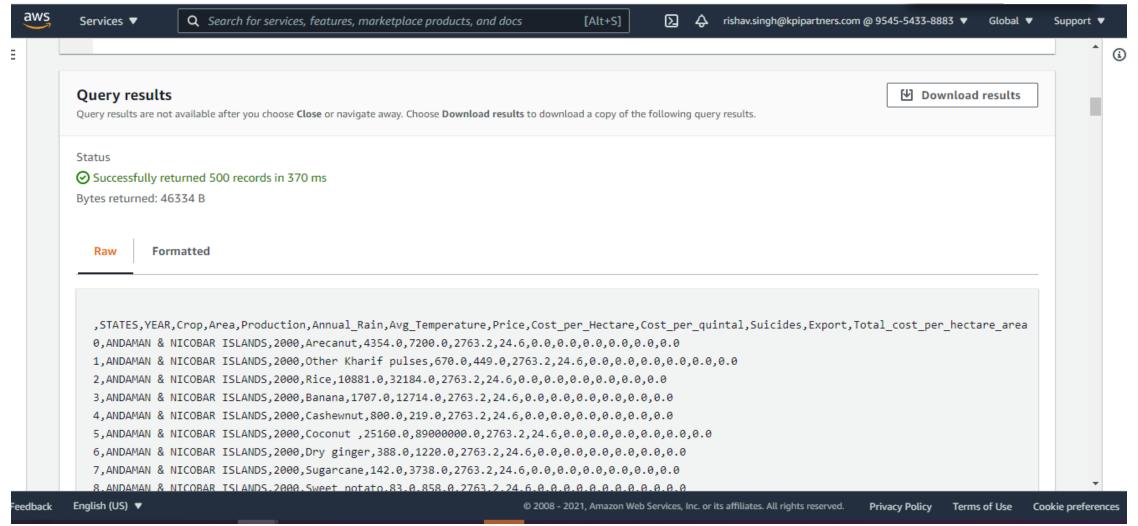


Cont...



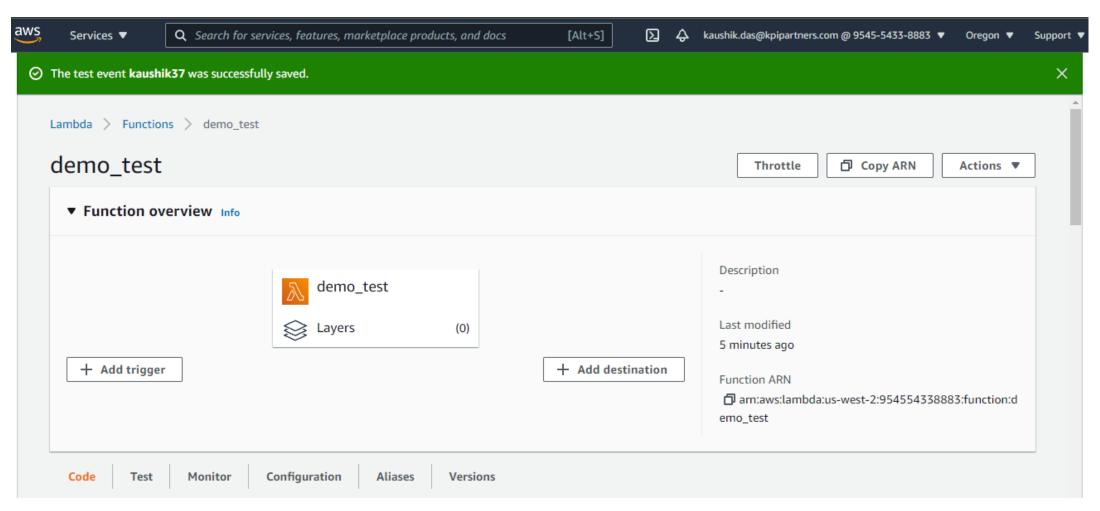






Retrieving data from URL using Lambda Function





Cont...



