Data Analytics Final Project: A Look Into Space Launches around the World

Pedro Montano

Agenda

- Introduction
- Sourced Data
- Data Cleaning Log
- Hypothesis
- Data Query and Visualization
- Conclusion

Introduction

In this presentation we will be looking into data for space launches from 1957 (the first launch being the Sputnik) through July 2020. We were interested in providing insights into what this data shows us on the space programs in different countries.

Sourced Data

The data we are using was sourced through Kaggle (https://www.kaggle.com/datasets/salmane/space-missions-launches). As with most data out there, it needed a lot of cleaning as seen below. We will be going over our cleaning process in the next two slides.

	A	B	C	D	E	F	G	H	I
1		Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_Statu
2	0	0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	Fri Aug 07, 2020 05:12 UTC	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50	Success
3	1	1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Center, China	Thu Aug 06, 2020 04:01 UTC	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29.75	Success
4	2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive		Success
5	3	3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	Thu Jul 30, 2020 21:25 UTC	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65	Success
6	4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145	Success
7	5	5	CASC	LC-9, Taiyuan Satellite Launch Center, China	Sat Jul 25, 2020 03:13 UTC	Long March 4B | Ziyuan-3 03, Apocalypse-10 & N	StatusActive	64.68	Success
8	6	6	Roscosmos	Site 31/6, Baikonur Cosmodrome, Kazakhstan	Thu Jul 23, 2020 14:26 UTC	Soyuz 2.1a | Progress MS-15	StatusActive	48.5	Success
9	7	7	CASC	LC-101, Wenchang Satellite Launch Center, China	Thu Jul 23, 2020 04:41 UTC	Long March 5 | Tianwen-1	StatusActive		Success
10	8	8	SpaceX	SLC-40, Cape Canaveral AFS, Florida, USA	Mon Jul 20, 2020 21:30 UTC	Falcon 9 Block 5 | ANASIS-II	StatusActive	50	Success
11	9	9	JAXA	LA-Y1, Tanegashima Space Center, Japan	Sun Jul 19, 2020 21:58 UTC	H-IIA 202 | Hope Mars Mission	StatusActive	90	Success
12	10	10	Northrop	LP-0B, Wallops Flight Facility, Virginia, USA	Wed Jul 15, 2020 13:46 UTC	Minotaur IV | NROL-129	StatusActive	46	Success
13	11	11	ExPace	Site 95, Jiuquan Satellite Launch Center, China	Fri Jul 10, 2020 04:17 UTC	Kuaizhou 11 | Jilin-1 02E, CentiSpace-1 S2	StatusActive	28.3	Failure
14	12	12	CASC	LC-3, Xichang Satellite Launch Center, China	Thu Jul 09, 2020 12:11 UTC	Long March 3B/E | Apstar-6D	StatusActive	29.15	Success
15	13	13	IAI	Pad 1, Palmachim Airbase, Israel	Mon Jul 06, 2020 01:00 UTC	Shavit-2 | Ofek-16	StatusActive		Success
16	14	14	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Center, China	Sat Jul 04, 2020 23:44 UTC	Long March 2D | Shiyan-6 02	StatusActive	29.75	Success
17	15	15	Rocket Lab	Rocket Lab LC-1A, M?Âhia Peninsula, New Zealand	Sat Jul 04, 2020 21:19 UTC	Electron/Curie | Pics Or It Didn??Â¦t Happen	StatusActive	7.5	Failure

Data Cleaning Procedure

- Separated Location column into “State” and “Country” using Text to Column Function
- Moved countries from state column and marked “N/A” for cells where a state does not apply.
- Marked “N/A” under “Country” column for cells where no specific country was listed.
- Removed second column which was a duplicate of the first column.
- Renamed first column to “Index.”
- Renumbered line number column starting from 1 instead of 0.
- Removed UTC and Day of the week short forms to change the data types to date and time respectively.
- Separated “Date and Time” column into “Date” and “Time (UTC)” columns.
- Separated the “Details” Column into “Rocket” and “Mission_and_Payload” columns.

Data Cleaning Procedure

- Removed redundant “status” in front of all the “statusactive” and statusretired” entries in “Rocket_Status” to make it easier to read.
- Changed the title of “price” column to include the units (million USD).
- Removed all corrupted characters that arose from special characters in other languages and the “degree” or “No.” symbol (º). Researched launches for accuracy of the “Organization,” “Rocket,” “Mission and Payload,” “Site_Name,” and “State” columns.
- Converted “Mission_Status” to 0 and 1 values instead of the four text options.
- Converted the “Index” column to “launch_id” instead. This column has the sequence of 1 to 4324 starting from the oldest launch instead of the newest.

Clean Data: So Clean Mr. Clean Approves

	A	B	C	D	E	F	G	H	I	J	K	L
1	launch_	Organisati	Date	Time (UT	Rocket	Mission_and_Payloa	Rocket_Stat	Price (Million US	Mission_Stat	Site_Name	State	Country
2	4324	SpaceX	7-Aug-20	5:12	Falcon 9 Block 5	Starlink V1 L9 & Black	Active	50	1	LC-39A Kennedy Space C	Florida	USA
3	4323	CASC	6-Aug-20	4:01	Long March 2D	Gaofen-9 04 & Q-SAT	Active	29.75	1	Site 9401 (SLS-2) Jiuquan	N/A	China
4	4322	SpaceX	4-Aug-20	23:57	Starship Prototype	150 Meter Hop	Active	NULL	1	Pad A Boca Chica	Texas	USA
5	4321	Roscosmos	30-Jul-20	21:25	Proton-M/Briz-M	Ekspress-80 & Ekspres	Active	65	1	Site 200/39 Baikonur Cos	N/A	Kazakhstan
6	4320	ULA	30-Jul-20	11:50	Atlas V 541	Perseverance	Active	145	1	SLC-41 Cape Canaveral A	Florida	USA
7	4319	CASC	25-Jul-20	3:13	Long March 4B	Ziyuan-3 03, Apocalyp	Active	64.68	1	LC-9 Taiyuan Satellite La	N/A	China
8	4318	Roscosmos	23-Jul-20	14:26	Soyuz 2.1a	Progress MS-15	Active	48.5	1	Site 31/6 Baikonur Cosm	N/A	Kazakhstan
9	4317	CASC	23-Jul-20	4:41	Long March 5	Tianwen-1	Active	NULL	1	LC-101 Wenchang Satelli	N/A	China
10	4316	SpaceX	20-Jul-20	21:30	Falcon 9 Block 5	ANASIS-II	Active	50	1	SLC-40 Cape Canaveral A	Florida	USA
11	4315	JAXA	19-Jul-20	21:58	H-IIA 202	Hope Mars Mission	Active	90	1	LA-Y1 Tanegashima Spac	N/A	Japan
12	4314	Northrop	15-Jul-20	13:46	Minotaur IV	NROL-129	Active	46	1	LP-0B Wallops Flight Faci	Virginia	USA
13	4313	ExPace	10-Jul-20	4:17	Kuaizhou 11	Jilin-1 02E, CentiSpace	Active	28.3	0	Site 95 Jiuquan Satellite L	N/A	China
14	4312	CASC	9-Jul-20	12:11	Long March 3B/E	Apstar-6D	Active	29.15	1	LC-3 Xichang Satellite La	N/A	China
15	4311	IAI	6-Jul-20	1:00	Shavit-2	Ofek-16	Active	NULL	1	Pad 1 Palmachim Airbase	N/A	Israel
16	4310	CASC	4-Jul-20	23:44	Long March 2D	Shiyan-6 02	Active	29.75	1	Site 9401 (SLS-2) Jiuquan	N/A	China
17	4309	Rocket Lab	4-Jul-20	21:19	Electron/Curie	Pics Or It Didn't Happe	Active	7.5	0	Rocket Lab LC-1A Mahia	N/A	New Zealand

Hypothesis

After our initial review of the data we came up with the following hypothesis:

“While the USA consistently maintains the highest number of launches since 1959, they also maintain the highest failed launch percentage.”

Data Query on mySQL

1. What are the top 5 countries with the highest number of launches (excluding the USA)?

- Russia, Kazakhstan, France, China, and Japan.

Total, Succesful, and Failed Launches

Country	Total Laun..	Successful Laun..	Failed Launch ..
Russia	1,398	1,305	93
USA	1,349	1,186	163
Kazakhstan	701	608	93
France	303	285	18
China	269	244	25
Japan	126	113	13
India	76	63	13
N/A	36	33	3
New Zealand	13	11	2
Iran	13	4	9
Israel	11	9	2
Kenya	9	9	0
Australia	6	3	3
North Korea	5	2	3
South Korea	3	1	2
Brazil	3	0	3
Spain	2	2	0
Iran	1	1	0

Failed Launch Count, Successful Launch Count and Total Launch Count broken down by Country.

Data Query on mySQL

2. What is the percentage of failed launches for the USA and the next 5 countries from question No. 1?
   - USA Failed/Total Launches= 163/1,349= 12.08%
   - Russia Failed/Total Launches= 93/1,398= 6.65%
   - Kazakhstan Failed/Total Launches= 93/701= 0.1326= 13.27%
   - France Failed/Total Launches= 18/303= 0.0594= 5.94%
   - China Failed/Total Launches= 25/269= 0.0965= 9.29%
   - Japan Failed/Total Launches= 13/126= 0.1092= 10.31%

3. Which is the country with the lowest percentage of failed launches?
   - France

4. Comparing the USA launches to these 5 countries; do they have a lower percentage of failed launches or relatively the same?
   - Only Japan and Kazakhstan have percentage higher than 10, but the other 3 are below 10%. France, Russia, and China have the lowest percentages and as such are more successful with their launches.

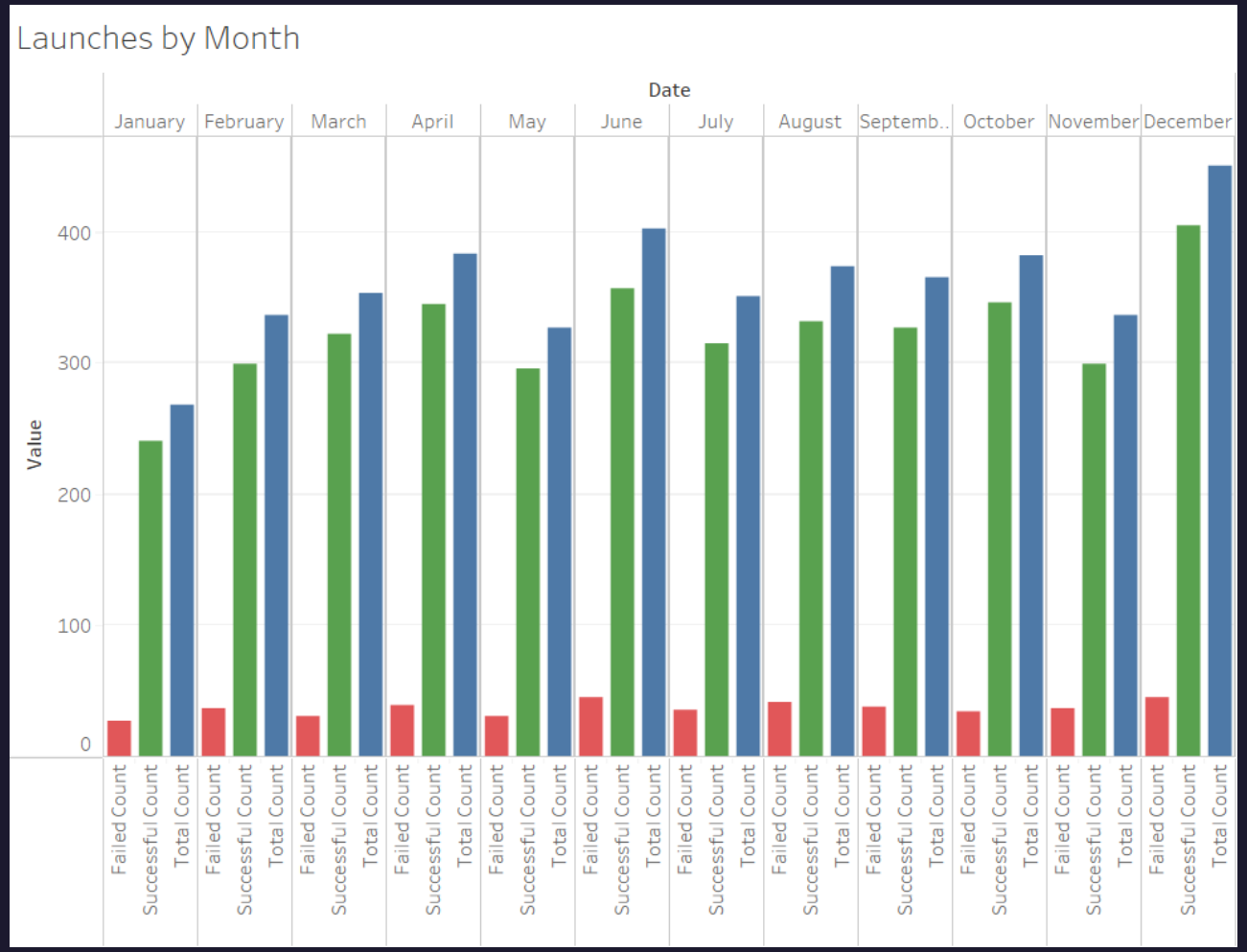
Percentage of Failed Launches

Country	
Brazil	100.000000%
Iran	69.230769%
South Korea	66.666667%
North Korea	60.000000%
Australia	50.000000%
Israel	18.181818%
India	17.105263%
New Zealand	15.384615%
Kazakhstan	13.266762%
USA	12.083024%
Japan	10.317460%
China	9.293680%
N/A	8.333333%
Russia	6.652361%
France	5.940594%
Spain	0.000000%
Kenya	0.000000%

Data Query on mySQL

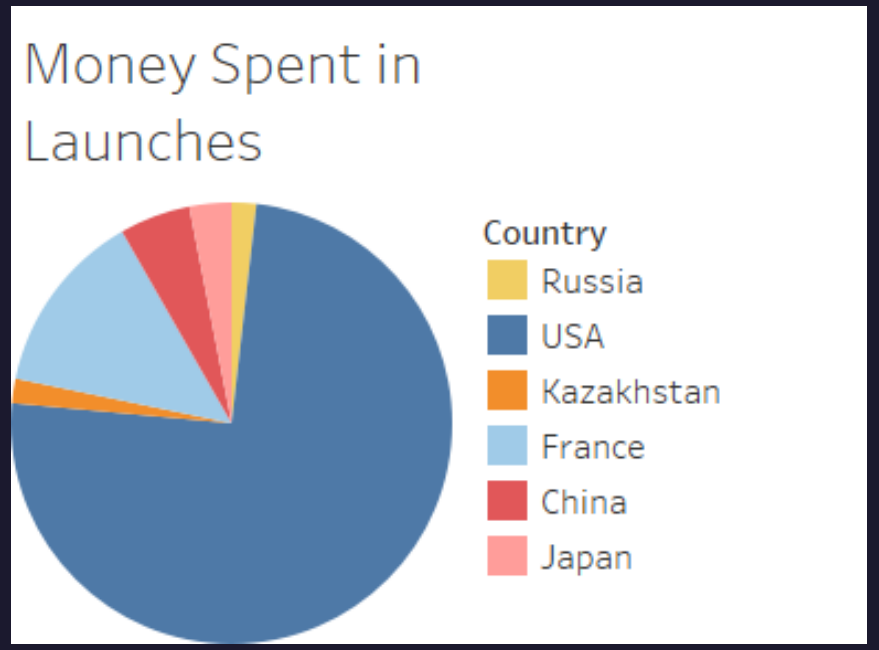
5. Is there a correlation between the month of the launch and a successful or failed launch?

- The most successful and failed launches are in June or December. This seems to be because most launches are held in those two months.



Data Query on mySQL

6. What was the costliest failed launch and which country was it? (Consider inflation)
    - The most expensive failed launch cost $450M in 1986, The Challenger.



Conclusion

Russia has the highest total launches while they hold the second lowest failed launch rate in the top five countries with most total launches. Additionally, Kazakhstan has the highest failed launch rate within those same top five countries. It is important to note that Russia and Kazakhstan were both part of the USSR and yet more launches were still taking part in Russia before the dissolution of the USSR in 1991. It is possible we may not have enough data on the money spent on launches from Russia, but they do beat the US in terms of launch count.