An Analysis of Historical Flood Claims Data

An analysis of FEMA historical flood claims was performed using Apache Spark and Scala. The data was obtained from the OpenFEMA datasets found at the following links:

https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1

https://www.fema.gov/openfema-data-page/fima-nfip-redacted-policies-v1

The data was used to answer the following questions:

1. Which 3 states have the highest and lowest claim (# of claims) to policy (# of policies) ratios? Consider policyCount field.

States with the Lowest Claim to Policy Ratio

Property State	Policy Counts	Claim Counts	Claim to Policy Ratio
NM	28543	1425	0.05
ID	12116	1126	0.09
NV	23430	2273	0.10

States with the Highest Claim to Policy Ratio

Property State	Policy Counts	Claim Counts	Claim to Policy Ratio
PR	17241	28115	1.63
VI	2919	4636	1.59
AL	55945	86476	1.55

All of the states with a low claim to policy ratio are relatively dry states, while those with a high claim to policy ratio are coastal.

2. What is the average total (building and contents) policy, average premium (total policy cost), and average claim amount for each occupancy type and the ratio of claim payout to policy amount and the ratio of claim payout to premium amount.

(1 = single family residence; 2 = 2 to 4 unit residential building; 3 = residential building with more than 4 units; 4 = Non-residential building; 6 = Non-Residential Business)

Policy, Premium and Claim Amounts by Occupancy Type

Occupancy Type	Average Policy Amount	Average Premium	Average Claim Payout	Ratio of Claim to Total Policy Amount	Ratio of Claim to Policy Premium
4	393511.04	2330.94	47307.41	0.12	20.30
3	1221096.77	2567.01	65743.36	0.05	25.61
2	286211.57	1170.92	30650.54	0.11	26.18
6	423769.73	2692.81	95443.64	0.23	35.44
1	262711.86	761.20	35089.66	0.13	46.10

Single family residences have the largest claim payout to policy premium ratio. Non-Residential businesses have the largest claim payout to total policy amount ratio.

3. Which zip codes had more than 5000 claims in any given year (ignoring policy count)

Claim County Zip Code and Year of Loss

Zip Code	Year of Loss	Claim Count
70065	2005	10783
08008	2012	9894
70122	2005	9638
70043	2005	8546
70003	2005	8443
70126	2005	8092
70124	2005	7551
70458	2005	7375
70119	2005	6936
11561	2012	6551
70072	2005	6218
70117	2005	5886
70127	2005	5540
70058	2005	5527
70001	2005	5514
70128	2005	5482
77550	2008	5410
08226	2012	5365

All of these zip codes correspond to locations hit by one of three major hurricanes in the corresponding years:

2005, Hurricane Katrina- Louisiana
2008, Hurricane Ike- Texas
2012, Hurricane Sandy- New York/ New Jersey

4. Which city had the most loss for each state for each decade?

Cities with Largest Total Loss By Decade

Decade	City	State	Total Loss ($)*
1980s	Houston	TX	136,540,492
1990s	New Orleans	LA	227,341,928
2010s	Houston	TX	4,498,711,000
2000s	New Orleans	LA	6,877,078,312

**Not adjusted for inflation*

In terms of total loss, for each decade since 1980, either Houston, TX or New Orleans, LA has been hit the hardest by floods.