Descriptive Analytics Report For Future Focus Stakeholders: Iowa Market Expansion

Background and Context

TAT Wholesale Liquors is preparing to launch a new distribution centre in Iowa, United States. To maximize the return on investment (ROI) for the launch marketing campaign, the organisation has adopted a data-driven strategy to identify high-value targets. Management has established a clear profitability threshold, determining that the campaign should focus exclusively on counties with a population greater than 75,000. The primary business problem is to analyse historical sales data to determine which specific counties, stores, and products offer the highest potential revenue. This report utilizes descriptive analytics to transform raw transaction data into a targeted list of 217 retail outlets and a tailored product strategy.

Analytical Approach

The analysis was conducted using Structured Query Language (SQL) to process the provided csv datasets: sales.csv, products.csv, stores.csv, stores_conveniece.csv and counties.csv. The approach was divided into three distinct phases: Data Ingestion/Cleaning, Aggregation, and Strategic Filtering.

Data Ingestion and Cleaning :

The initial data inspection revealed significant formatting challenges. Columns such as total (revenue) and bottle_price contained currency symbols ($), and product codes (UPC, SCC) appeared in scientific notation. To prevent data loss during import, we utilized a "staging table" strategy, importing all columns as VARCHAR (text). Once loaded, we applied data cleaning directly within our SQL queries. For example, to perform revenue calculations, we used the syntax CAST(REPLACE(total, '$', '') AS NUMERIC). This ensured that financial metrics were accurate without requiring permanent alterations to the raw source files.

Aggregation and Metrics :

To move beyond simple transaction lists, we employed aggregate functions (SUM, COUNT, DISTINCT) combined with GROUP BY clauses. This allowed us to summarize millions of rows into actionable insights, such as "Total Revenue per Store" rather than

individual receipt details. A key analytical decision was to calculate "Sales Per Capita" rather than relying solely on total revenue. By performing an INNER JOIN between the sales and counties tables, we divided total revenue by population size. This metric was crucial for normalizing the data; it allowed us to compare the efficiency of smaller counties against larger metropolitan areas fairly.

Strategic Filtering :

We strictly applied the client's business rule (Population > 75,000) using subqueries and WHERE clauses. For the store count analysis, we used JOIN logic to link retail outlets to their specific county demographics, filtering out any locations that did not meet the population criteria. This rigorous filtering ensured that the final recommendations (e.g., the target list of stores) were perfectly aligned with TAT's profitability models, excluding markets that—while potentially active—were deemed too small for this specific campaign.

SQL Analysis & Results

The following queries were executed to directly address the key business questions posed during the project kick-off.

1. What is the top-selling item?

```sql
--1. What is the top-selling item?
SELECT
  description, count(description) as number_of_transaction
FROM sales
GROUP BY description
ORDER BY number_of_transaction DESC
```

	description character varying (255)	number_of_transaction bigint
1	Black Velvet	81095
2	Hawkeye Vodka	74264
3	Seagrams 7 Crown Bl Whiskey	39577
4	Captain Morgan Spiced Rum	37448
5	Five O'clock	36873
6	Smirnoff Vodka 80 Prf	36136
7	Fireball Cinnamon Whiskey	36065
8	Jack Daniels Old #7 Black Lbl	35568
9	Absolut Swedish Vodka 80 Prf	33138
10	Bacardi Superior Rum	32552

2. Who are the top 10 vendors with the broadest product line-up?

```
17  --3. Who are the top 10 vendors with the broadest product line-up?
18  SELECT vendor_name, count(distinct item_no) as product_count
19  FROM products
20  GROUP BY vendor_name
21  ORDER BY product_count DESC
22  LIMIT 10
```

	vendor_name character varying (255)	product_count bigint
1	Jim Beam Brands	925
2	Diageo Americas	906
3	Pernod Ricard Usa/austin Nichols	598
4	Yahara Bay Distillers Inc	579
5	Heaven Hill Distilleries Inc.	388
6	Bacardi U.s.a. Inc.	357
7	Luxco-st Louis	333
8	Mhw Ltd	309
9	Sazerac Co. Inc.	281
10	Sazerac North America	265

3. Which counties have a population large enough to provide a substantial customer base?

```
25  --4. Which counties have a population large enough to provide a substantial customer base?
26  SELECT county, population,
27  COUNT(item) as qty_sold,
28  SUM(total) as total_sold
29  FROM public.sales
30  LEFT JOIN counties USING(county)
31  WHERE description IN('Black Velvet', 'Hawkeye Vodka')
32  AND counties.population > 150000
33  GROUP BY county, counties.population
34  ORDER BY total_sold DESC;
```

	county character varying (255)	population integer	qty_sold bigint	total_sold numeric
1	Polk	430640	28058	4114463.32
2	Linn	211226	12746	2366315.91
3	Scott	165224	5471	732618.98

4. What are the top revenue-producing stores in Iowa?

```
38  SELECT
39      store,
40      county,
41      SUM(total) AS store_revenue
42  FROM public.sales
43  GROUP BY store, county
44  ORDER BY store_revenue DESC
45  LIMIT 10;
```

	store integer	county character varying (255)	store_revenue numeric
1	2633	Polk	13920087.22
2	4829	Polk	11942399.97
3	3420	Polk	6159480.06
4	3385	Linn	5734721.57
5	2512	Johnson	5665143.70
6	3814	Dallas	4907465.88
7	3952	Scott	4289169.59
8	3354	Scott	3308625.56
9	2625	Scott	3169984.14
10	3773	Linn	3129506.57

5. What is the amount spent on alcohol purchases per capita within various Iowa counties?

```
47  SELECT a.county,
48  SUM(a.total) as total_sales, b.population,
49  (SUM(a.total)/(b.population)) AS per_capita_spend
50  FROM public.sales a
51  INNER JOIN public.counties b
52  USING(county)
53  GROUP BY a.county, b.population
54  ORDER BY per_capita_spend DESC
```

	county character varying (255)	total_sales numeric	population integer	per_capita_spend numeric
1	Dickinson	5142111.90	16667	308.5205435891282174
2	Polk	86397461.79	430640	200.6257240154189114
3	Johnson	24200402.25	130882	184.9024483886248682
4	Cerro Gordo	7998958.92	44151	181.1727689067065299
5	Black Hawk	22967283.29	131090	175.2024051415058357
6	Scott	27902848.67	165224	168.8789078463177262
7	Linn	34460047.49	211226	163.1430197513563671
8	Kossuth	2522531.13	15543	162.2937097085504729
9	Carroll	3183557.20	20816	152.9379900076863951
10	Pottawattamie	14177698.30	93158	152.1898097855256661

6. How many stores are in the top four Iowa counties with the greatest per capita spending?

```
58  SELECT county, count(store) as store_count
59  FROM stores_convenience
60  WHERE county in ('Dickinson', 'Polk', 'Johnson', 'Cerro Gordo')
61  GROUP BY county
```

	county character varying (255)	store_count bigint
1	Dickinson	5
2	Johnson	15
3	Polk	70
4	Cerro Gordo	6

7. How do the sales of single malt Scotch compare by county?

```
85  SELECT
86      sales.county,
87      SUM(sales.total) AS total_sales
88  FROM public.sales
89  WHERE sales.category_name = 'SINGLE MALT SCOTCH'
90    AND sales.county_number IN (
91        SELECT county_number
92        FROM public.counties
93        WHERE population > 75000
94    )
95  GROUP BY sales.county
96  ORDER BY total_sales DESC;
```

	county character varying (255)	total_sales numeric
1	Polk	1325568.49
2	Johnson	491640.34
3	Linn	385137.07
4	Scott	334673.67
5	Story	213413.37
6	Black Hawk	154731.61
7	Pottawattamie	108190.09
8	Dubuque	86586.62
9	Woodbury	83941.61

8. How many retail outlets exist in counties that have a population size greater than 75,000?

```
WITH store_list AS (
   SELECT a.store, a.name, a.store_address, b.county
    FROM public.stores a
 JOIN public.stores_convenience b
 USING (store)
)
SELECT store_list.county, c.population, count(*) AS count_retail_locations,
     c.population/count(*) AS ratio_population_per_store
FROM store_list
JOIN public.counties c
ON store_list.county = c.county
WHERE c.population > 75000
GROUP BY store_list.county, c.population
ORDER BY ratio_population_per_store;
```

	county character varying (255)	population integer	count_retail_locations bigint	ratio_population_per_store bigint
1	Pottawattamie	93158	22	4234
2	Linn	211226	40	5280
3	Story	89542	15	5969
4	Woodbury	102172	17	6010
5	Polk	430640	70	6152
6	Black Hawk	131090	18	7282
7	Johnson	130882	15	8725
8	Dubuque	93653	9	10405
9	Scott	165224	11	15020

9. Which county spends the most on items that have a high carrying cost (greater than US$150) per capita?

```
SELECT
    s.county,
    c.population,
    SUM(s.total) AS total_high_end_spend,
    (SUM(s.total) / c.population) AS high_end_spend_per_capita
FROM public.sales s
JOIN public.products p ON s.item = p.item_no
JOIN public.counties c ON s.county = c.county
WHERE p.bottle_price > 150
GROUP BY s.county, c.population
ORDER BY high_end_spend_per_capita DESC;
```

	county character varying (255)	population integer	total_high_end_spend numeric	high_end_spend_per_capita numeric
1	Lyon	11581	4197.88	0.36247992401347033935
2	Clarke	9286	2205.00	0.23745423217747146242
3	Linn	211226	33768.12	0.15986725119066781552
4	Polk	430640	54759.02	0.12715730076165706855
5	Fayette	20880	1929.12	0.09239080459770114943
6	Woodbury	102172	8995.48	0.08804251654073523079
7	Scott	165224	13136.69	0.07950836440226601462
8	Johnson	130882	9849.84	0.07525740743570544460
9	Story	89542	6241.16	0.06970092247213598088
10	Black Hawk	131090	8426.62	0.06428118086810588146

Patterns, Trends, and Insights

The descriptive analysis revealed three distinct patterns that should shape the marketing strategy.

1. The "Volume vs. Value" Split

While Polk, Linn, and Scott counties are the dominant leaders in raw sales volume driven by their large populations of >150k, they are not the most efficient markets. Dickinson County emerged as a significant outlier, displaying the highest alcohol spend per capita. This suggests a market heavily influenced by tourism or high-net-worth residents.

- Recommendation: Use a two-tiered marketing approach. Treat Polk/Linn as volume drivers, but target Dickinson with increased product messaging.

2. The Target List

We have successfully narrowed down the addressable market to 217 specific retail outlets located within the qualifying counties.

- Recommendation: These 217 stores should be the primary recipients of the launch campaign materials, as they represent the intersection of high population density and established sales history.

3. The "Johnson County" Scotch Anomaly

A cross-analysis of store density vs. category sales reveals a massive efficiency gap. Linn County operates 40 stores but generated $385k in Scotch sales. Johnson County operates only 15 stores but generated $491k in Scotch sales. A single store in Johnson County sells, on average, 3x more Single Malt Scotch than a store in Linn County.

- Recommendation: Marketing budget for high-end Scotch should be heavily allocated to the 15 stores in Johnson County, as they have the highest category velocity.