**Logo, company name

Description automatically generated**

**Assessing Retail Sales for an Anonymous Company and The External Factors That Affect It**

**Technical Report**

**Prepared By:**

Abdullah Majali 20200378

Tamara Otaibi 20200961

Abdelrahman Elayyan 20200441

Dana Twal 20200959

**Course Name**: Data Visualisation **Course Instructor:** Prof. Moh’d Azzeh

**1. Abstract**

This study aims to explore the external factors that might affect sales at a chain of stores across the country. This data was collected at interesting places in each store ranging from the stock keeping units (SKUs) at the back of the store to the cash registers at the front. Management is interested in finding key insights that can help maximize profits and store revenue at each of the recorded branches. They also hope to compare their brick-and-mortar sales vs online-store sales. Our investigation into the sales data has led us to a few conclusions: Customers are most willing to spend the most on holidays and when promotions are available, and capitalizing on these factors should in theory bring about more sales revenue to the stores.

**2. Introduction**

**2.1 Overview**

There are several factors that a business may need to consider to be profitable business, such as attracting customers with catchy promotions and fulfilling their needs and purchases while also keeping track of their employees’ performances. We conducted an exploratory data analysis of the dataset provided, then pre-processed and engineered new features so that it can be ready for various visualizations and tests that can help better understand the data and yield solutions for the problem at hand.

**2.2 Dataset**

The dataset originally consists of 4000 records and 33 features with 5 date features: customer birthdate, employee birthdate, employee hire date, order date, and promotion date. Four continuous numerical features: Price, Rate, city area, and population density. Three discrete numerical features: Quantity, Working Hours, and City Population. Fifteen categorical features: Customer first name and last name, employee first name and last names, customer email, employee email, product category, department, branch location, branch name, employee type, promotion type, payment type, city name, and employee role description. And finally 3 binary features: gender, holiday, and customer type.

**3. Methodology**

**3.1 Target selection**

Since our research is centered around studying factors affecting our stores' revenues, we will conclude our results from the price column. This column describes the total amount spent at check out by each customer. We can group the dataset by different features to isolate how these factors affect our sales.

**3.2 General EDA**

Our data exploration approach was to check the data types we are working with and If there are any missing/inconsistent records in our dataset. The data set contained no missing or inconsistent values so we moved on to formulating new more useful features to aid our analysis.

**3.3 Feature Engineering**

We were able to extract new features from our dataset that might help provide clearer insight for our analysis. Customer/Employee birthdates were used to calculate their ages, employee hire dates were used to calculate employee tenure, order dates were used to determine sales quarters, and customer city was used to group our data by region.

**3.4 Outliers**

With the help of box plots, we were able to spot potential outliers in Quantity, City\_Area, and population columns, we also suspected some in other columns so we decided to use Rosner’s test due to its suitability to the problem at hand, as the data is larger than 25 data points. The results of Rosner's test are as follows:

| **Variable** | **Value** | **Obs.Num** | **Outlier** |
| --- | --- | --- | --- |
| Quantity | 180 | 312 | FALSE |
| Price | 999.9 | 578 | FALSE |
| Price | 998.6 | 2961 | FALSE |
| Price | 998.5 | 3203 | FALSE |
| City\_Population | 2473400 | 30 | FALSE |
| City\_Area | 32832 | 17 | FALSE |
| City\_Area | 32832 | 22 | FALSE |
| City\_Density | 723.4 | 2 | FALSE |
| City\_Density | 723.4 | 10 | FALSE |
| City\_Density | 723.4 | 11 | FALSE |
| City\_Density | 723.4 | 50 | FALSE |
| City\_Density | 723.4 | 55 | FALSE |

It was concluded that there were no outliers so we elected to leave these values as is since they did not pose an obstacle to our analysis.

**3.5 Findings**

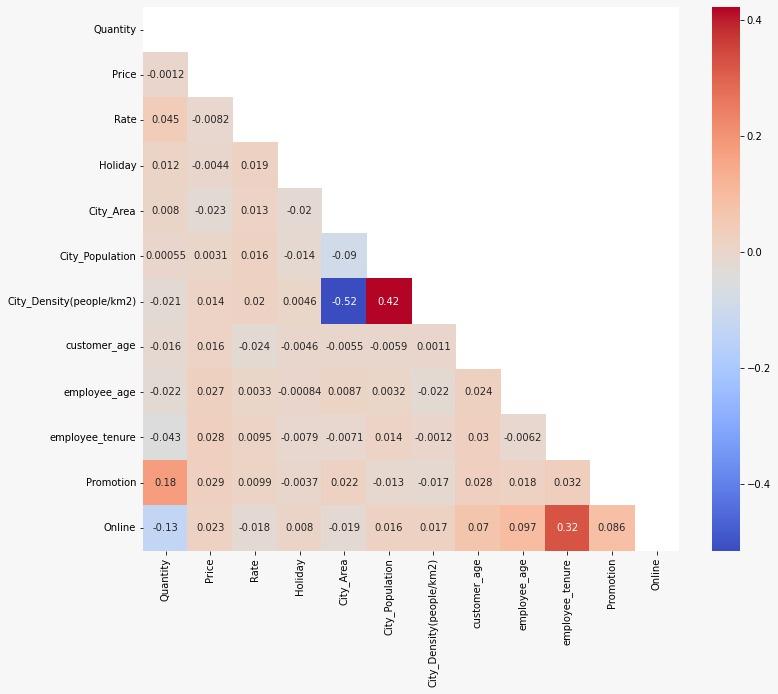
Our findings were centered around a few main features that seemed to affect store revenue. Mostly promotions, holidays, customer membership, customer age group, and employee tenure. The latter 2 features were derived from customer birthdates and employee hire dates, respectively.

**4. Results**

According to our analysis, higher revenues were accredited to both customers and employees. we saw that on average customers were willing to spend more on holidays and promotions; memberships also had an effect since members have spent more overall in the store than non-members. The majority of the sales also came from adult customers and that can be attributed to the fact that customers aged 20-59 had more money to spend than retirees ( 60+) and teenagers (11-20). In addition, employees with the longest tenures tended to complete more sales orders and were handling a majority of online orders which netted higher sales than their less experienced colleagues.

**6.  Appendix**

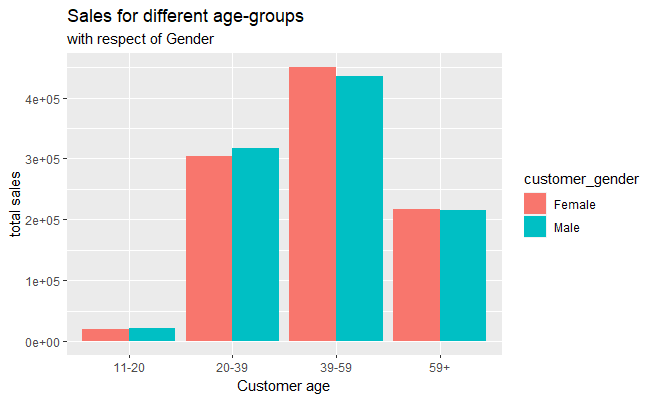
**6.1 Figures**

******

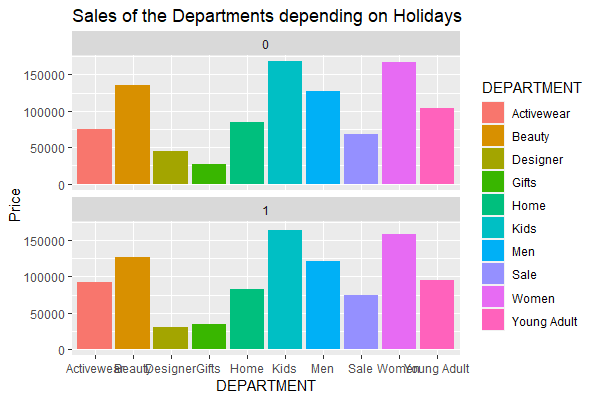
*Fig 6.1.1 Correlation heatmap*

******

*Fig 6.1.1 outlier detection using boxplots*

**

*Fig 6.1.3 sales collected from customer age groups*

**

*Fig 6.1.4 customer spending per department with respect to holidays*