

# **Reduce the Wait Time For Customers at Checkout**

**Predict the Number of Small Basket Sizes On A Given Day**

**Group 6A:**

**Pankaj Sharma - 61310346**

**Bhaskar Kandukuri – 61310697**

**Varun Unnikrishnan – 61310181**

**Santosh Gowda – 61310163**

**Anuj Bajpai - 61310663**

# Business Problem

- **Objective:**

- Reduce the waiting time for customers at check out

- **Challenges:**

- Limited Number of Total Available Checkout Lanes
- Uncertainty about basket size distribution on a given day

- **Approach:**

- Dynamically alter the percentage of “Fast” checkout lanes based on demand predictions
- Balance customer density between “Fast” and “Regular” Checkout lanes

# Data Mining Problem

- **Objective:**

- To predict the % of Small Baskets on a given day

- **Definitions:**

- Small Basket: Baskets with a total item quantity less than 20
- Total Number of Check out Lanes: 50

- **Predicted Variable:**

- Number of “Fast” Checkout lanes on a given day

- **Predictors:**

- Number of “Fast” check out lanes over the last 10 Days - Seasonality
- Number of “Fast” check out lanes over the last 4 weeks on that day – Weekly Trends

# Data

Used 1 Year's worth of daily basket level data

**Step – 1:** For each day, calculate the average basket size, std deviation of basket size.

**Step – 2:** For each day, calculate the % of small baskets using a threshold of 20 (Assuming a Normal Distribution of Quantity)

**Step – 3:** Calculate the number of “Fast” Checkout Lane

**Step – 4:** For each day, calculate the average number of “Fast” checkout lanes over the last 10 days (Predictor 1) as well as on the same day for the last month (Predictor 2)

# Benchmark

Simple Linear Average of % of Small Baskets on that day for the past year

## Methods

- KNN
- Regression

## Snapshot

Transaction Date	Avg(Total Quantity)	StdDev(Total Quantity)	Percentile of TQ	check out lines	Day	Predictor 1	Predictor 2
8/1/2011	71.44	86.38	27.58%	15	2	15	15
8/2/2011	58.9	69.07	28.67%	7	3	11	11
8/3/2011	61.55	68.39	27.17%	6	4	10	10

# Evaluation

To Evaluate our models, we compared the actual number of “fast” checkout lanes in the test data versus the predicted number of “fast” checkout lanes using our models

- **Actual Vs Benchmark:**

Average Difference = 1.24 Fast Checkout Lanes

- **Actual Vs Regression:**

Average Difference = 1.105 Fast Checkout Lanes

- **Actual Vs KNN:**

Average Difference = 1.16 Fast Checkout Lanes

## Insights

Based on the results from our data mining exercise, we found that the Regression model gave us more accurate predictions. However, this was the case because the amount of data available was very less. In case there is more data available, the KNN model would work much better and should be used in lieu of the Regression model.

**Thank You**