

Advance analytics
internship coding
challenge sai charan



**ASSIGN
BUSTER**

The number of distinct website visits; 1 session may have multiple visits
distinct_sessions | The number of distinct website visitors; 1 session may
have ultimate visits orders | The number of website orders gross_sales | The
total gross sales for website orders bounces | The number of visits that only
viewed one page add_to_cart | The number of visits that added a product to
cart product_page_views | The number of product pages viewed
search_page_views 1 The number of search pages viewed Exploratory data
analysis First we explore relevant data, company site visited by users is
described by next table: Acme 7392 Booty Pinnacle 5725 Shortly 5532
Tabular Widget Acme site, result to be the more visited followed by Pinnacle
and Shortly.

In platforms according to next table, the most visitors users use ISO,
followed by Android devices data of database origin. 410 pad 459 Other 327
Android 3172 phone Symbiosis 74 BlackBerry 1 589 Linux 2036 Unknown
1641 Chromes ISO 1349 3435 Macintosh Macros 333 2054 Windows
Windowpanes 2399 1315 Worth analyses the general direction of sales on
gross sales sites gross_sales variable represents how they were developing
sales per day at each site; in the following figure shows the aggregate sales
all sites. Linear regression model could bring an estimate of how this months
has developed. Graphic review explains that last months bring a possible
cycle effect resulting to be the best for selling.

Previous plot also shows that best sales overcome in the last month of the
year. This could bring business opportunities, stock planning among others
factors for years to come. Previous figure, shows the gross sales for all
websites. Results can be understood by comparing both graphics that gross
<https://assignbuster.com/advance-analytics-internship-coding-challenge-sai-charan/>

sales are very close related to the amount of sales that has been taken in 2013. To ensure this graphic assumption here is a regression analysis. ## call: lm(formula = sales\$gross_sales ~ ordered\$orders) Residuals: Min IQ Median -74332 -40590 -6319 Max 35378 118136 Coefficients: (Intercept) ordered\$orders --Signing. Codes: 4291 . 274 10077. 800 0. 426 0. 671 143. 36 1 . 966 72. 822 <2e-16 *** o 0. 001 0. 01 0. 05 0. 1 Residual standard error: 48350 on 266 degrees of freedom Multiple R-squared: 0. 9522, Adjusted R-squared: 0. 9521 F-statistic: 5303 on 1 and 266 DE, p-value: < 2. 2+16 Results show that exist a high correlation between the amount of orders and the gross sales , in summary 95% of sales are explained by the amount of orders with a 9. 99% of accuracy. Next plot shows the sales prediction with ARIMA prediction. The brief of the analysis only permitted an autoarima forecasting method. Arima forecast modeling shows an stable growth for first 20 days of 2014. Possible causes of sales development.

At this point we have seen sales over time development and a simple regression analysis, next is need to see which variables possible has effects on sales. For that next is developed a correlation matrix analysis to explore the possible correlations that exist between variables a possible explanations. Correlation matrix are shown graphically (See annex for full results). Blue represents positive and significant correlation, as blue is darker means that this relationship is stronger. Lighter blue means week relationship. White means no relationship. By a brief review we can see next assumptions. - Gross sales are highly influenced by orders and add cart (regression analysis confirmed it). Visits are highly influenced by bounces, this means a high amount of visitors Just visit one page. However also visits

are highly explained by - Visits are also explained in a different way; distinct sessions explain a good amount of visitors, that means visitors are generally diverse and different. Conclusions and possible future analysis. Analysis has been through different stages. Initially, the general sales overview of 2013 has shown that sales exhibit a cycle effect. To ensure this as an assumption, we require to analyze data from other years. Correlation analysis has shown a high and almost total relationship between orders and sales, this explains that most of orders are not cancelled or affected by a return policy.

January of 2014, still this analysis had been reviewed a more punctual forecasting loud be required. Third, has been identified all users preferences for ISO and Android devices. Gross sales has 2 possible explanations: orders and add to cart options. Further analysis. I would recommend a more punctual analysis for each variable that require special treatment, exploratory analysis gives a general and important picture of the scenario of the companies. However I would suggest next possible projects. *

Specialized forecasting with time series for prediction. * Performance for each company. * Analysis for market segmentation of devices. Annex

Correlation matrix results by significance levels.