**Software Requirements Specification**

**for**

**<Project>**

**Version 1.0 approved**

**Prepared by :**

**Sweta Jaiswal**

**Tayibah**

**Sharda university**

**22/11/2018**

**Table of Contents**

**Table of Contents** [**ii**](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.3whwml4)

**Revision History** [**ii**](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.2bn6wsx)

**1. Introduction** [**1**](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.qsh70q)

1.1 Purpose [1](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.3as4poj)

1.2 Document Conventions [1](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.1pxezwc)

1.3 Intended Audience and Reading Suggestions [1](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.49x2ik5)

1.4 Product Scope [1](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.2p2csry)

1.5 References [1](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.147n2zr)

**2. Overall Description** [**2**](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.3o7alnk)

2.1 Product Perspective [2](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.23ckvvd)

2.2 Product Functions

2.3 Operating Environment

**3. External Interface Requirements** [**3**](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.32hioqz)

3.1 User Interfaces [3](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.lnxbz9)

3.2 Hardware Interfaces [3](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.1hmsyys)

3.3 Software Interfaces [3](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.41mghml)

3.4 Communications Interfaces [3](https://docs.google.com/document/d/10kVDFkL_N99J_uKOr25to6gS3GLSYUyTKhyAEb_sRPM/edit#heading=h.2grqrue)

**4. Further**

4.1 How to make a Time Series Stationary

4.2 Forecasting of time series

4.3 Components of time series

1. **Introduction**
   1. **Purpose**

Since my first week on this platform, I have been fascinated by the topic of time series analysis. This kernel is prepared to be a container of many broad topics in the field of time series analysis. My motive is to make this the ultimate reference to time series analysis for beginners and experienced people alike.

* 1. **Document Conventions**

This document follows MLA Format. Bold-faced text has been used to emphasize section and sub-section headings. Highlighting is to point out words in the glossary and italicized text is used table and recognize diagrams

* 1. **Intended Audience and Reading Suggestions**

This document is to be read by the development team, the project managers, marketing staff, testers and documentation writers. Our stakeholders, company manufacturing associated hardware, company providing embedded operating system, shareholders of Project and distributors who markets the finished product, may review the document to learn about the project and to understand the requirements. The SRS has been organized approximately in order of increasing specificity. The developers and project managers need to become intimately familiar with the SRS.

* 1. **Product Scope**

Time series analysis involves developing models that best capture or describe an observed time series in order to understand the underlying causes. This often involves making assumptions about the form of the data and decomposing the time series into constitution components. The objective of time series is to develop a mathematical model and then estimate the model to predict future patterns.

* 1. **References**
* **“The Analysis of Time Series: An Introduction” by C Chatfield**
* **“Time Series Analysis Forecasting and Control” by G E P Box and G M Jenkins**
* **“Time Series Analysis” by James Douglas Hamilton**

1. **Overall Description**
   1. **Product Perspective**

This dataset lends itself to a some very interesting visualizations. One can look at simple things like how prices change over time, graph an compare multiple stocks at once, or generate and graph new metrics from the data provided. From these data informative stock stats such as volatility and moving averages can be easily calculated. The million dollar question is: can you develop a model that can beat the market and allow you to make statistically informed trades!

* 1. **Product Functions**

As the name suggests, TS is a collection of data points collected at constant time intervals. These are analyzed to determine the long term trend so as to forecast the future or perform some other form of analysis. But what makes a TS different from say a regular regression problem? There are 2 things:

1.It is time dependent. So the basic assumption of a linear regression model that the observations are independent doesn’t hold in this case.

2.Along with an increasing or decreasing trend, most TS have some form of seasonality trends, i.e. variations specific to a particular time frame. For example, if you see the sales of a woolen jacket over time, you will invariably find higher sales in winter seasons.

* 1. **Operating Environment**

To be used efficiently, all computer software needs certain hardware components or other software resources to be present on a computer. These prerequisites are known as (computer) system requirements and are often used as a guideline as opposed to an absolute rule. Most software defines two sets of system requirements: minimum and recommended. With increasing demand for higher processing power and resources in newer versions of software, system requirements tend to increase over time. Industry analysts suggest that this trend plays a bigger part in driving upgrades to existing computer systems than technological advancements.

Software requirements specification establishes the basis for an agreement between customers and contractors or suppliers on how the software product should function (in a market-driven project, these roles may be played by the marketing and development divisions). The software requirements specification lays out functional and non-functional requirements, and it may include a set of use cases that describe user interactions that the software must provide.

1. **External Interface Requirements**
   1. **User Interfaces**

The external users are the clients. All the clients can use this software for indexing and searching.

* 1. **Hardware Interfaces**

The external hardware interface used for indexing and searching is personal computers of the clients. The PC’s may be laptops with wireless LAN as the internet connections provided will be wireless.

* 1. **Software Interfaces**

The Operating Systems can be any version of Windows.

* 1. **Communications Interfaces**

The Operating Systems can be any version of Windows**.**

**4. How to make a Time Series Stationary?**

Though stationarity assumption is taken in many TS models, almost none of practical time series are stationary. So statisticians have figured out ways to make series stationary, which we’ll discuss now. Actually, its almost impossible to make a series perfectly stationary, but we try to take it as close as possible.

Lets understand what is making a TS non-stationary. There are 2 major reasons behind non-stationaruty of a TS:

1. Trend – varying mean over time. For eg, in this case we saw that on average, the number of passengers was growing over time.

2. Seasonality – variations at specific time-frames. eg people might have a tendency to buy cars in a particular month because of pay increment or festivals.

5. **Forecasting a Time Series**

We saw different techniques and all of them worked reasonably well for making the TS stationary. Lets make model on the TS after differencing as it is a very popular technique. Also, its relatively easier to add noise and seasonality back into predicted residuals in this case. Having performed the trend and seasonality estimation techniques, there can be two situations:

1- A strictly stationary series with no dependence among the values. This is the easy case wherein we can model the residuals as white noise. But this is very rare.

2- A series with significant dependence among values. In this case we need to use some statistical models like ARIMA to forecast the data.

Our project is based on ‘Time Series Analysis’ where the main focus areas are:

1. Understanding Time Series
2. Data Exploration
3. Time Series Forecasting using different methods

**Understanding Time Series**

A time series is a series of [data points](https://en.wikipedia.org/wiki/Data_point) indexed or listed or graphed in time order. Most commonly, a time series is a [sequence](https://en.wikipedia.org/wiki/Sequence) taken at successive equally spaced points in time. Thus it is a sequence of [discrete-time](https://en.wikipedia.org/wiki/Discrete-time) data.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time series forecasting is the use of a [model](https://en.wikipedia.org/wiki/Model_(abstract)) to predict future values based on the previously observed values.

Time series are very frequently plotted via [line charts](https://en.wikipedia.org/wiki/Line_chart). Time series are used in [statistics](https://en.wikipedia.org/wiki/Statistics), [signal processing](https://en.wikipedia.org/wiki/Signal_processing), [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), [mathematical finance](https://en.wikipedia.org/wiki/Mathematical_finance), [weather forecasting](https://en.wikipedia.org/wiki/Weather_forecasting), [earthquake prediction](https://en.wikipedia.org/wiki/Earthquake_prediction), [astronomy](https://en.wikipedia.org/wiki/Astronomy), [communications engineering](https://en.wikipedia.org/wiki/Communications_engineering), and largely in any domain of applied [science](https://en.wikipedia.org/wiki/Applied_science) and [engineering](https://en.wikipedia.org/wiki/Engineering) which involves [temporal](https://en.wikipedia.org/wiki/Time) measurements.

While [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

**Components of a Time Series**

* Trend : Trend is a general direction in which something is developing or changing. It may be an increasing trend, decreasing trend or constant in this time series.
* **Seasonality** : Seasonality is when the pattern is repeating at regular time interval. It is any predictable change or pattern in a time series that recurs or repeats over a specific time period.

### Problem Statement

Unicorn Investors wants to make an investment in a new form of transportation - JetRail. JetRail uses Jet propulsion technology to run rails and move people at a high speed! The investment would only make sense, if they can get more than 1 Million monthly users with in next 18 months. In order to help Unicorn Ventures in their decision, we need to forecast the traffic on JetRail for the next 7 months.

### Here we need to first understand the data. Understanding of data includes Hypothesis Generation, Getting the system ready and loading the data, Dataset Structure and Content, Feature Extraction, Exploratory Analysis.

After that comes the forecasting using Multiple Modeling Techniques. For this we need to first split the data into training and validation part, then apply various modeling techniques.

Various Modeling techniques that we will work upon for forecasting the time series are :

**i) Naive Approach**

* In this forecasting technique, we assume that the next expected point is equal to the last observed point. So we can expect a straight horizontal line as the prediction.

**ii) Moving Average**

* In this technique we will take the average of the passenger counts for last few time periods only.

**iii) Simple Exponential Smoothing**

* In this technique, we assign larger weights to more recent observations than to observations from the distant past.

**iv) Holt’s Linear Trend Model**

* It is an extension of simple exponential smoothing to allow forecasting of data with a trend. This method takes into account the trend of the dataset. The forecast function in this method is a function of level and trend.

**v) SARIMAX model on daily time series**

* SARIMAX model takes into account the seasonality of the time series. Order in the above model represents the order of the autoregressive model, the degree of differencing(number of times the data have had past values subtracted) and the order of moving average model.