Chapter 3

Research Methodology and Research Design

This chapter discusses the research methodology adopted for this study. Further, the chapter narrates the population, sampling method and the sample size of this study. It explains the results of the pilot study. Thereafter, process of data collection, characteristics of the sample, data cleaning process, and data distribution are explained. Moreover it explains scale of variables, and measures adopted for this research endeavor. In addition, it has investigated the reliability and validity study of the constructs. Finally, the measurement model results are presented.

Research Stance

A phenomenological study explains the meaning for lived experiences of several individuals with the context of a concept or an object or a phenomenon with the direct experience. Phenomenologists focus on describing what all individual participants have in common as they experience the same concept or object or phenomenon (Creswell 2007). The basic purpose of phenomenology is to converge participant experiences with a phenomenon to a description of the shared common essence (a “grasp of the very nature of the thing,” van Manen, 1990).

In this study, pragmatism was used to explore the IT Services startup phenomenology which is intended to observe the technology startup phenomena and also gain insights from the lived experiences of entrepreneurs who have been part of the research population. Along with pragmatism, interpretivism is adopted to attempt a generic critical success factor framework to be derived from the observed phenomenology and to assess the startup success framework.

A qualitative research method was well-suited to the current research study. The qualitative study helped the researcher to probes the participants to find deeper insight during interviews to understand the reasons behind various business challenges that lead to the failure of technology startups and if these challenges are overcome, the startup sustains and achieves business growth. The use of an in-depth interview data collection and analysis approach helped to facilitate a vivid description and deeper insights of the participants’ lived experiences and perceptions (Strong, Pyle, Devries, Johnston, & Foskett, 2008). The current
phenomenological study used purposive/judgmental sampling to identify the population. Snowball sampling method was used to select the specific sampling data from the references received from the participants and validate with the sampling criteria from the population of entrepreneurs/ founding team members from Information Technology services startups based out of India.

**Research Method**

The use of semi-structured one-on-one interviews as the major data collection method starts with the assumption that the participants’ perspectives are honest, meaningful, and that their perspectives have an impact the success of information technology services startups and the current study. One of the two types of interviews, in-depth interview is used in this research study to encourage free and open responses and in-depth exploration of limited set of questions as a questionnaire.

Qualitative research work focuses typically in depth on relatively small samples, even single case or data, selected purposefully in comparison with quantitative data which are typically selected randomly on large samples. (Patton, 2002). The techniques used in quantitative and qualitative methods of sampling are different as the purpose of each strategy is different (Patton, 2002).

The realistic outcomes from purposeful sampling lies in selecting information-rich samples for studying in depth. The goal of purposeful sampling is to select information-rich sampling data or cases whose study illuminates the questions under study (Patton, 2002). There are several different strategies for purposefully selecting information-rich cases, in our research work, identifying and selecting entrepreneurs / senior leaders , led the researcher to a knowledgebase which can be mined to get meaningful knowledge from purposeful selection of samples. This approach helped to get insightful deep resourceful information which highlights the answers to the research objectives.

The maximum variation sampling method was used to design the research for sampling for data collection across technology startup concentrated clusters in **NCR region, North India and Bangalore region, South India**. Maximum Variation sampling strategy for purposeful
sampling helped in capturing principal outcomes that cut across participant concentration in the identified clusters (Patton, 2002).

The maximum variation sampling strategy helped to turn the weakness of small sample size into strength. Any common theme that emerged from variation on various aspects are of value in capturing the core experiences and common shared aspects (Patton, 2002).

The data collection and analysis of diversified small sample has two kinds of findings: (1) Descriptions of each aspect, which are resourceful for identifying unique and core themes, and (2) Commonly identified shared patterns found across diversified participants and derive their importance from emerging out of heterogeneity (Patton, 2002). The researcher studied information that are significant common patterns within the identified variation in sampling strategy adopted(Patton, 2002).

**Research Design**

The methodology adopted in the current study is based on one-on-one interviews with 30 entrepreneurs/founding members/ senior management team members from Information Technology services companies, in cities across India as well as based on Information collection and analysis from supporting entrepreneurial ecosystem. The government of India and state startup initiatives and policies with details of tax exemptions, reimbursements, concessions are studied from official government websites and used this information to probe the participants on impact of those external ecosystem factors.

A phenomenological study explains the lived experiences of many individuals with a context of a concept or an object or a phenomenon. A phenomenology provides a deep understanding of a phenomenon as experienced by several individuals. The participants in the study are chosen to be individuals who have experienced the phenomena under study, so that the researcher, in the end, can forge a common understanding. Bracketing (Epoche) personal experiences was applied by the researcher to become separated from the text and apply interpretivism on the startup phenomena under study. Phenomenology can include only single interview or multiple interviews with participants for data collection (Creswell, 2007).
The development of ideal-types from the lived experiences shared a paradigm which gives objective view of a phenomenon experienced by the participants under study.

The methodology adopted in this research study is based on one-on-one interviews with 30 entrepreneurs/Senior Management members from Information Technology services companies started in various cities across India. This is substantiated by the support ecosystem provided by government bodies and semi-government bodies via initiatives and policies and the participants’ view on the external support system.

**Defining Population**

Information Technology services startup companies incubated in India during 2000-2012 which have service portfolio in areas of – Application (Software) Development and Maintenance Services, Infrastructure Management Services, Testing Services, IT Consulting and Training, System Integration Services, Cloud Computing Services, Data Analytics Services and Mobile Computing Services. The founders/entrepreneurs/senior management team members who have been part of the above Information Technology Services Startups form the population for the current research study.

**Sampling**

Sample data for current research work is identified from reliable databases like Information Technology services registered with NASSCOM, ROC and CII to prepare the sample data for **IT services startups during 2000 – 2012.**

NASSCOM database had 1467 registered IT companies. After filtering out the sample criteria, the database had potential 431 privately held IT Services Startups incorporated during 2000-2012. **Judgmental/Purposive sampling** is used to identify the sampling data pertaining to the current study population as the study is exploratory and needs more probing on the lived experiences of entrepreneurs/Senior Management team members and need their complete involvement. **Snowball sampling** methods in particular to get to the intended sampling population for in-depth study via references and connections made with the participants who shared their views during the data collection.
Entrepreneurs / Senior Management team members who have been part of information technology services startup ventured in India during 2000-2012 in Information Technology services form the sample data. India, as one of the major Information Technology hub has IT services startups in Bangalore, Hyderabad, Chennai, Mumbai, Delhi and Pune.

Sampling Distribution

The geographical distribution of business incubations across India was not uniform: 56% of incubators are present in southern India, 21% are located northern India, 17% are located western India, and only 6% are present in eastern regions of India (NSTEDB, 2014).

After analyzing the concentration of the industry in the above cities and startups concentration in general, the sample was taken in proportionate to the population of Industry hub and startup ecosystem locations, majorly having sampling data from Bengaluru, which is the only city in India listed in top 20 startup ecosystem ranking (Startup Genome Report, 2015).

Figure 10: Sample Population for the current research study
If we take region-wise across India, the sample distribution is as follows:

![Bar chart showing sample distribution across regions](image)

**Figure 11: Sample Population for the current research study (Region)**

**Data Collection**

**Mode/Method of Data Collection:**

- Primary data is collected via personal interviews – face-to-face or telephonic, based on pre-determined, semi-structured domain based questionnaire to drive in-depth interviews to explore the entire 360 degree experiences in all perspectives of a startup ecosystem.
- Secondary data is collected specifically on government startup initiatives and policies and probed participants to understand their views on these support initiatives during their startup journey.
- Secondary data is collected from official websites, various articles, citations on the Startup companies that the participants had their work experiences in the selected
Information Technology Services startups and this data is used to validate the primary data collected during interviews.

**Interview Techniques**

The use of the personal interviews by the researcher alone (face-face and telephonic) helped to facilitate validity and reliability. Audio taping of interviews was done to review any information shared during interview and transcription of the recorded data took place for further data analysis. Triangulation was conducted to achieve internal validity by using mixed methods (formal interviews, casual interactions/discussions on same entities) and the information already collected via secondary data collection methods. The interview periods ranged from 50 minutes to 1 hour 45 minutes with the participating individuals. Since the interview was based on a semi-structured pre-prepared questionnaire (Annexure D), it made interview data extraction easier into the research constructs via various “categories” or “parts” of interview sections used to find answers to the current research questions.

**Pilot Study**

Conducting a pilot study gives early warning about where the core research objective possibly has flaws or whether proposed research methods or instruments are inappropriate or too complicated (David De Vaus, 1993). A pilot study can be the pre-testing of a particular research instrument (Baker, 1994). Researcher began with "qualitative data collection and analysis on a relatively unexplored topic, using the results to design a subsequent quantitative
phase of the study" (Tashakkori & Teddlie, 1998). Welman and Kruger (1999) proposed the following three values of a pilot study:

“It is needed to detect possible flaws in measurement procedures (including instructions, time limits, etc.)

In the operationalisation of independent variables. A pilot study is also valuable to identify unclear or ambiguous items in a questionnaire.”

The current study used pilot study to validate the items in the questionnaire that was needed to progress to data collection. The non-verbal behaviour of participating individuals in the pilot study helped in identifying the discomfort experienced concerning the wording or order of content or question formulation in a questionnaire which can be rectified and used in further data collection.

In the current research work, initial draft of semi structured questionnaire was used to drive the interview with 3 pilot participants identified and interviewed. The flaws in the order of questions, irrelevant questions, discomforting questions which needed participants to share financial parameters of the startups during interview were removed / corrected. With these changes to the questionnaire, participants concentrated on giving the interview with confidence and knowledge sharing mindset. Participants shared their experiences along with the advices that they would like to give to a new entrant into the Information Technology services sector, even though they have taken a few actions which they admitted as their mistakes or wrong decisions, which they were willing to share as their learnings and give advice to avoid or mitigate those decisions/paths taken during the study.

**Validity and Reliability**

Validity and reliability are “important in establishing the truthfulness, credibility, or believability of findings” (Neuman, 2003). The data collection procedure, data analysis methods, and reporting of findings were reliable and valid to ensure the validity of the current research study. The research design of the study helped to mitigate the threats to internal validity and external validity.
Reliability

The reliability of a study is a measure of the consistency of the research findings. The reliability in the current study occurred through the adoption of Moustakas’s (1994) systematic techniques of data analysis procedures. Proper organization of the data with regard to themes or sections and the handling of categories helped to improve the reliability of the current study (Richards, 2006).

Validity

The validity of a qualitative study is a reflection of the accuracy and credibility of the outcomes (Neuman, 2003). The validity of a study indicates the stability, credibility, trustworthiness, and consistency of the findings (Cooper & Schindler, 2003). The validity of a research study is a measurement of “the extent to which measurements achieve the purpose for which they are designed” (Simon, 2006). The current study included two types of validity: internal validity and external validity.

Internal validity

The internal validity of the current qualitative study was “the extent to which its design and the data that it yields allow the researcher to draw accurate conclusions about correlations and relationships within the data” (Leedy & Ormrod, 2001, p. 103). The main intent of maintaining internal validity is to reduce internal errors that may exist in design methodology (Neuman, 2003). Internal validity is an accurate reflection of the phenomenon under study and a measurement of the accuracy and credibility of the findings (Leedy & Ormrod, 2005).

The adoption of practices such as data triangulation, external independent audit helped to ensure the internal validity of the current study (Cooper & Schindler, 2003). An examination of outcomes from all data sources to gather evidence supporting a theme helped to achieve triangulation (Creswell, 2007). The transcriptions of audio taped data, researcher’s field notes, and participants’ body language/ tone of the voice comprised the basis for triangulation (Creswell, 2007). An external individual reviewed the audio recordings and several other aspects of the research for accuracy of the transcriptions which also provides internal validity.
The use of epoché, or bracketing techniques helped to eradicate threats to the validity and further support the internal validity of the current study (Groenewald, 2004). Moustakas (1994) says, “The challenge of the ‘epoché’ is to be transparent to ourselves, to allow whatever is before us in a consciousness to disclose itself so that we may see with new eyes in a naive and completely open manner”. The current study was free of influence from the researcher’s preconceived ideas, perceptions or judgments.

**External validity**

External validity “is the extent to which its results apply to situations beyond the study itself—in other words, the extent to which the conclusion drawn can be generalized to other contexts” (Leedy &Ormrod, 2001). External validity is a measure of achieving meaningful coherence between results, data, and the techniques and approach by which the results are found (Garza 2007). The purpose of the current qualitative study was not to generalize, but to explore the lived experiences of participants regarding success factors for Information Technology Services start-up phenomenon. The current study included a limitation on generalization of the findings to other populations.

**Data cleaning**

There was no necessity to remove any case of sample data as data was based on interviews, even though some of the questions weren’t answered by the participants and may not add to significant findings. During data analysis, the non-responsiveness for any particular question is noted as outliers as appeared in Cluster membership formation and considered before arriving at conclusions.

**Data Analysis**

The research study is qualitative in nature and the perspectives of participants are captured as participants conveyed it. The entire interview data, along with researcher’s field notes with audio transcription of interviews is entered into Google Form template mimicking the semi-structured questionnaire (Annexure D) to help analysis of data.

Since the data collected is semi-structured in nature, while in-depth perspectives are captured via interviews, conscious questions designed to capture the intensity or level of agreement, Likert’s rating scale is used for particular interview questions during the interview. A Likert
scale is a psychometric scale developed by psychologist, Rensis Likert in 1932, commonly involved in research that employs questionnaires where scaling responses for questions are needed for data collection.

The format of five-level Likert scaling response item used in the current research’s open ended questions to participants are:

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Closed questions with rating scale options</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>

Likert scaling is a bipolar scaling method, which measures either positive or negative response to a statement. When responding to a Likert scaling item, participants specify their level of agreement or disagreement on agree-disagree scale for questions. The range of the scale captures the intensity of participants’ feelings for a given item.

The above closed question is followed by open question “to share the perspectives”, which is followed by an advisory question “to share their advice to upcoming entrepreneurs”. This is the typical textual responses which are recorded and transcribed.

Google form used as the interim software to hold the data collected was later downloaded after the completion of interviews as an excel data download. Since the record set is 30 and qualitative nature of the study makes excel – charting and summarizing capabilities is sufficient to bring out data analysis of the current research study.

In Deductive analysis, data is analyzed according to an existing framework (Patton 2002). Inductive analysis involves discovering patterns, themes, and categories in one’s data (Patton 2002). In the current mixed method research study, both inductive and deductive analysis are used which are applied at specific data collection and analysis processes. In qualitative analysis, inductive analysis takes place in early stages when developing a codebook for content analysis, often called as open coding (Patton 2002). Grounded theory helps researcher becoming immersed in the data, being grounded so that embedded meanings and relationships can emerge (Patton 2002). The final confirmatory stage of qualitative analysis is
deductive analysis and affirming the appropriateness of the inductive content analysis (Patton 2002).

According to Moustakas (1994), the steps or procedures involved in data analysis of a phenomenological study includes horizontalizing, creation of themes, and development of textural and structural descriptions regarding the phenomenon. Horizontalizing is the process where every statement or horizon relevant to the topic is treated as equally important (Moustakas, 1994). From the horizontalized statements, the meaning units are listed and clustered in to themes followed by development of the textural descriptions of the experience. Content Analysis usually refers to analyzing text in interview transcripts, diaries or documents (Patton 2002). Finally, the essences of the phenomenon are constructed from emerging textural descriptions and structural descriptions.

**Figure 13:** Data Analysis Process
The first step of data analysis process included the data preparation, including data entry, editing, coding, and conversion of raw data into reduced and classified forms that are more appropriate for analysis (Cooper & Schindler, 2003). Transcription of the recorded data into text occurred. To protect anonymity, and insure confidentiality, participants’ names were replaced with number codes. Care was taken to insure that the names of the participants and their organizations did not appear in the data. To ensure the confidentiality of participants, transcribed data received different file names; and the names of organizations and persons referred by participants were masked (Conklin, 2007).

The qualitative data analysis process commenced following the completion of horizontalization. Qualitative data analysis is an iterative process of individual and group level review and interpretation. The data analysis portion of qualitative study “consists of obtaining a general sense of the data, memoing ideas, thinking about the organization of the data, and considering whether more data are needed” (Creswell, 2007). The next step was the development of themes with the intent to find answers to the research questions and further explore the successful start-up phenomenon.

First step of data analysis is classifying data or coding the data collected (Patton 2002). Coding is a qualitative data analysis process that includes sorting, categorizing, and organizing the meaning units of collected data (Jones, 2007).

A code structure is created which is a compilation of the emerging codes. Brief descriptions or definitions for each code along with any guidance for when / how to use the codes is also developed. Every line of interview transcript is coded under coding structure using an integrated method of purely inductive grounded method and start list method.

---

**Figure 14: Coding**
Purely inductive grounded method used participant experiences to drive analysis, develops code de novo line by line, and limits possibility of researcher forcing a preconceived result. Start list method uses preliminary organizing framework where initial codes are drawn from researcher expertise, topic of enquiry and existing literature. This is somewhat deductive in nature. An integrated approach retains the benefits of inductive coding, begins with broad code types and then develop sub code types from data.

Content Analysis is used to refer to any qualitative data reduction and sense making effort that takes a volume of qualitative data and tries to identify core consistencies and useful meanings (Patton 2002). Content Analysis sometimes refers to searching text for repeating words or themes (Patton 2002). In the current research, word counts from the interview data to identify repetitive or recurring words which emphasize specific selection of attributes for each homogenous variable are determined.

**Coding Structure for the current research Study**

Researcher derived a coding structure based on the semi structured interview questionnaire as abroad base. Various attributes were organized into logically fitting categories / types and were placed in A, B, C D and E parts of section in the semi-structured interview questionnaire.

<table>
<thead>
<tr>
<th>Broad code types at Level 1</th>
<th>Code sub types at Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part A:</td>
<td>Following attributes are identified under Part A as sub code types:</td>
</tr>
<tr>
<td>Information Technology</td>
<td>1. Type of Legal Entity</td>
</tr>
<tr>
<td>services Startup Profile</td>
<td>2. Year of establishment</td>
</tr>
</tbody>
</table>

Entrepreneur/Management Team member exited the
| 4. | Number of consultants/outsourcing partners (currently/When Entrepreneur/Management Team member exited the Venture) |
| 5. | Number of Locations of the Startup (sales locations including partnerships in different countries) |
| 6. | Number of Branches in India |
| 7. | Number of Branches (International) |
| 8. | Information Technology services offered |
| 9. | Type of EXIT Strategy (Planned if not Exited Yet) |
| 10. | Year of EXIT (If Exited) |
| 11. | Company Revenue (in Rupees, Crores) at the time of Exit/Interview |
| 12. | Customer Base (in Count) at the time of Exit/Interview |
| 13. | Profitability (in Rupees, Crores) at the time of Exit/Interview |
| 14. | Industry verticals served |

**Part B:**

| Information Technology services Startup company Investment Profile |
| Following attributes are identified under Part B as sub code types: |
| 1. Investment Profile |
| 2. Current Status of the Company |

**Part C:**

| Promoter/Founding Member / Senior Executive Management team member Profile |
| Following attributes are identified under Part C as sub code types: |
| 1. Gender |
| 2. Age |
| 3. Education |
| 4. Number of years of work Experience prior to Startups |
| 5. Role played |
| 6. Motivation |

**Part D:**

| Role of |
| Following attributes are identified under Part D as sub code types: |
| 1. Prior successful Startups |
|---------------------------------------------------------------|--------------------------------------------------------------------------------|

**Part E:**

Following attributes are identified under Part E as sub code types:

1. Organization Structure during INITIAL (First 2 years) phase
2. Organization Work Culture during INITIAL (First 2 years) phase
3. Organization Structures during GROWTH (Years 2-5) phase
4. Organization Work Culture during GROWTH (Years 2-5) phase
5. Founder/Promoter/Executive Management's Leadership/Managerial Style with the technology Startups

**Part F:**

Following attributes are identified under Part F as sub code types:

1. Market
2. Availability of Skilled Resources
3. Government Regulations
4. Government Tax policies

**Part G:**

Following attributes are identified under Part G as sub code types:

1. Partnerships
2. Add End-to-end solutions
3. Diversification
4. Newer/added Service Portfolio
5. Acquisitions
Table 3: Coding types and Sub Types for research constructs

Alignment of Code Structure with Research Questions

- PART A, PART B, PART C – forms the demographic data of the sample population, which gave the quantitative means of analyzing demographic data.

- Other parts - PART D, PART E, PART F, PART G, PART H, PART I, PART J of data structured at broad level, contribute to build various constructs which help answer the research questions.

Construct one: Startup team’s dynamics in Technology Services Startups

There is a need to explore and discuss the Information Technology services team dynamics which is essential for the business success of Information Technology services Startup in
India. There is a need to understand the team prior background plays a role in collaborative working and contributes to the startup success. Prior research has looked at some of the attributes in team dynamics (Chorev and Anderson, 2006), however there is a gap in research study on how these attributes play a role in Information Technology services companies in India. In this construct, relevant data pertaining to Information Technology services startup team’s prior background dynamics are grouped to be analyzed. This construct includes attributes or variables like prior successful startup work experience, prior failed startup experience, Level of Education, Marital Status, Work Experience, Age, Gender of founding members / senior management team members. Prior work on critical success factors for technology startups included team expertise (Chorev and Anderson, 2006). Prior research study has linked team expertise to technology business success (Chorev and Anderson, 2006). This construct determines the dimensions in team dynamics and its relation to the business challenges for Information Technology services startups in India whether they positively influence the success of startup business or negatively affect the startup business.


**R1: What are the founding team dynamics which contribute to success of IT Services startups in India?**
Construct 2: Leadership Styles at Information Technology services startups

Prior research has elucidated the link between entrepreneurship and leadership within organizations. Henning (2014) conducted a comparative study and found that overall impression is that Startups and larger organizations have a lot in common when it comes to leadership (Henning, 2014). Outstanding differences in their way of leading isn’t found between Startups and mature firms ((Henning, 2014). Zahra (2012) studied the correlation between different components of transformation and transactional leadership and companies’ performance in custom development. In this construct, dimensions of leadership styles followed or witnessed by participants is explored for Information Technology services startups in India.

Construct 3: Role of Business plan, Intellectual Property in Information Technology startups

Construct 3 is identified to explore the essence and impact of having strategic business plan before the startup is commercially set up. Strategic business plan are essentially a need for any funding needs for the initial set up or growth of the startup. This construct looked at participant’s view of business plan, the contents of the plan, duration, scope covering what is the plan and how is the entrepreneur is envisioning the startup and what is the blueprint of the
plan expected to be executed, frequency of updation of business plan due to changes in external or internal environment. Also, the perceptions of the participants that the technology services startups also considers the Intellectual property as a success factor for the startup is explored in the construct.

**Construct 4: External Startup Support System and Information Technology Services Startup Strategies**

In Construct 4, Technology services startup support system factors are explored from interview data of participants. The external ecosystem dimensions studied are market, government regulations, tax policies, availability of talent. Also, the secondary data collected for relevant factors for government regulations, initiatives, tax policies and exemptions are referred along with participant’s view to correlate the relationships.

Secondary data collected regarding government of India startup initiatives, specific state government schemes are discussed in detail available during 1991-2016. The applicability
and usage of government startup initiatives to technology services startups in India during 2000-2012 and share the perspectives of participants on the leverage the government policies.

Construct 5: Internal Strategies at Information Technology Services Startups

Construct 5 is created to explore internal business challenges faced by technology services startups and understand the various internal strategies planned and executed by the startups to overcome the challenges to sustain and grow their startups in the technology services sector. This encompasses all the internal dimensions that a technology services need to look at for their business management and move towards business growth. The various dimensions studied under internal strategy are execution strategies adopted including customer acquisition strategy, customer delivery strategy, use of systems and processes, growth strategies adopted, services / solutions strategies, Organizational work environment and management styles practiced are probed during the interviews with the participants.
Clustering of data

The objective of cluster analysis is to map observations to groups ("clusters"). With clustering, observation data within each group are similar to one another with respect to independent variables or attributes of study. The groups or clusters themselves stand apart from one another. In other words, the objective of clustering is to segregate the observation data into unique homogeneous groups. Cluster analysis is an exploratory tool meant to support the qualitative data analysis (Laura, 2015).

The basic steps in clustering data analysis is listed down below.
Figure 15: Steps in Cluster Analysis

Researcher had asked the participants to provide a detailed explanation about each open question, probing to find narrated detailed descriptions from their experiences on the various situations, contexts and responses to the situations so that lived experiences are captured in detail to support the possibility of identification of themes / patterns in groups / clusters during data analysis. Data driven codebook was built following grounded theory approach.

Post initial coding, a basic review of the response cases based on specific categories of observations (e.g., demographic data, market strategies, hiring strategies, service portfolio, and solution mix offered etc.) is performed. However, in depth interview data collected itself directly was not useful to identify trends within the data.
Cluster analysis techniques help segregate and organize sample cases into discrete groups, such that similarity is maximum within-group and similarity within group or cluster is minimized according to the independent variables considered for the cluster formation. Cluster analysis techniques operate on datasets for which well-defined or pre-specified, groups do not exist; Characteristics of attributes of independent variables are used to assign sampling entities into groups or clusters.

With cluster analysis, researcher can identify outliers and exclude them which are in small clusters and removed for further analyses. Cluster analysis assesses relationship even within a single set of variables using respondent data vs. defining relationship between independent and dependent variables. These groups or clusters which are formed based on single / multiple independent variables are studied for specific patterns of data redundancies so that similar entities or data domains can be interpreted as part of observations and findings summarized.

Clustering analysis techniques help the segregation and organization of multivariate data by grouping entities together into classes. Cluster analysis helps in ordering of the available data, making it sufficient for the analysis of small qualitative sample research data (Laura, 2015).

Clustering tools use databases with the observation data to be classified in rows, and the variables upon which the observation elements are to be classified are given in columns (Laura, 2015). Observation Data in qualitative research is often available as nominal, and in some cases they use ordinal variables (Laura, 2015). In nominal data, assigned numerical values represent strictly a name and are not intrinsically ordered (Laura, 2015). If ordinal data is present in qualitative data, assigned numerical values are ordered, but the differences between values cannot be quantified mathematically to derive meaningful information (Laura, 2015).

Clustering techniques organize data by comparing the values assigned for variables across classified rows of data, determining their level of similarity upon independent variables given in columns. Common distance measures are inappropriate with nominal or ordinal data (Laura, 2015). With nominal data in particular, which are not ordered, numerical values are by description or definition and not comparable to other numerical values quantitatively (Laura, 2015).
Hierarchical Clustering organizes similar entities or domains into classes or groups and arranges these classes or groups into a hierarchy. Hierarchical Clustering assesses and shows relationships expressed among the classified similar entities as groups. Primary use of non-hierarchical Clustering is to summarize data entities into fewer groups for subsequent data analysis which works well with large data sets. Non-hierarchical clustering has a limitation that it is not effective for elucidating relationships because there is no interesting structure within groups or clusters and no definition of relationships among clusters or groups is derived.

Polythetic Agglomerative Hierarchical Clustering (PAHC) techniques use the information on all the variables considered for cluster formations (i.e., polythetic). Each entity is initially assigned as a separate group or cluster. PAHC agglomerates each entity in a hierarchy of larger and larger groups or clusters until finally a single cluster contains all entities or domains.

There are many fusion algorithms which use different resemblance measures. Some of the main fusion strategies available in hierarchical clustering:

Single-Linkage (Nearest Neighbor): An entity's dissimilarity to a cluster is defined to be equal to its dissimilarity to the closest entity in that cluster; when two clusters agglomerate, their dissimilarity is equal to the smallest dissimilarity for any pair of entities with one in each cluster (Thomas and Pawel, 2006).

Complete-Linkage (Furthest Neighbor): An entity's dissimilarity to a cluster is defined to be equal to its dissimilarity to the furthest entity in that cluster; when two clusters agglomerate, their dissimilarity is equal to the greatest dissimilarity for any pair of entities with one in each cluster (Thomas and Pawel, 2006).

There are many other fusion algorithms like Centroid-Linkage (Unweighted Pair-Group Centroid), Median-Linkage (Weighted Pair-Group Centroid), Average-Linkage (Unweighted Pair-Group Average), Ward's Minimum-Variance-Linkage etc.

Fusion process in PAHC uses Nearest Neighbor--Euclidean Distance method. PAHC can be depicted using Tree-like plot called "Dendogram" depicting the agglomeration sequence in
which entities or cases are enumerated (identified) along one axis and the dissimilarity level at which each fusion of groups or clusters occurs on the other axis.

In the agglomerative hierarchical clustering method, in first step, each element is an individual cluster. All clusters are compared with those most similar, then merged to create a new cluster; this process repeats until a single cluster is formed. Final results present all the steps and easily visible in a Dendogram graph (Kaufman & Rousseeuw, 2009).

In current research study, "Average-Linkage" clustering method is used for clustering of data, which is a hierarchical agglomerative clustering method. "Hierarchical" because all clusters formed by these methods consist of mergers of previously formed clusters with similarity grouping depending on independent variables. "Agglomerative" because the methods begin with the same number of clusters as there are observations and end with a single cluster containing all observations. Since the sample size is 30, hierarchical clustering method was found to be opt for the current research study rather than non-hierarchical clustering method. Once hierarchical clustering is performed, the next step is to determine the number of clusters considered for analysis (Salvador & Chan, 2004).

**Cluster Validation**

The term clustering validation describes the process of evaluating the outputs of a clustering algorithm. Relative clustering validation, which evaluates the clustering structure by varying different parameter values for the same algorithm (e.g: varying the number of clusters k). It’s generally used for determining the optimal number of clusters.

**Internal clustering validation measures**

Internal clustering validation measures shows the compactness, the connectedness and separation of the cluster partitions formed.

1. Compactness measures evaluate how close the objects are within the same cluster. A lower within-cluster variation is an indicator of a good compactness (i.e., a good clustering). The different indices for evaluating the compactness of clusters are based
on distance measures such as the cluster-wise within average/median distances between observations.

2. Separation measures determine how a cluster is well separated from other clusters. The indices used as separation measures include:
   a. distances between cluster centers
   b. the pairwise minimum distances between objects in different clusters

3. Connectivity corresponds to the extent data rows are placed in the same cluster with the nearest neighbors in the observation cases.

**Silhouette analysis**

Silhouette analysis measures the clusters formed for the observations and it estimates the average distance between clusters formed as defined by cluster algorithm. This plot shows a measure of closeness from one point in one cluster to points in the neighboring clusters (Kaufman & Rousseeuw, 2009).

For each data row observation \(i\), the silhouette width \(s_i\) is calculated to determine is the number of clusters optimum for the dataset to be considered for data analysis and interpretation (Kaufman & Rousseeuw, 2009).

1. For each observation \(i\), calculate the average dissimilarity \(a_i\) between \(i\) and all other points of the cluster to which \(i\) belongs.
2. For all other clusters \(C\), to which \(i\) does not belong, calculate the average dissimilarity \(d(i,C)\) of \(i\) to all observations of \(C\). The smallest of these \(d(i,C)\) is defined as \(b_i=\min_C d(i,C)\). The value of \(b_i\) can be seen as the dissimilarity between \(i\) and its “neighbor” cluster, i.e., the nearest one to which it does not belong.
3. The **silhouette width** of the observation \(i\) is defined by the formula which is used in cluster validation:
   
   \[S_i = \frac{(b_i-a_i)}{\max(a_i,b_i)}.\]

**Interpretation of silhouette width**

Silhouette width is interpreted as follows:

1. Observations with a large \(s_i\) (almost nearing 1) are very well clustered
2. A small $s_i$ (nearing 0) means that the observation lies between two clusters
3. Observations with a negative $s_i$ are perhaps placed in the wrong cluster.

Once optimum number of clusters are formed, identification of elements being strongly present (or absent) was tested using Pearson’s chi square goodness of fit for the sample data to validate the analyzed results in the sample. A chi-square test of independence was performed to check the relation between cluster memberships and the variables considered for cluster classification. With the goodness of fit test results, the full qualitative data was revisited for the strong cluster group in the constructs to deduce more meaningful interpretation taking participant’s experiences and their startup/firm performance at the time of interview.

**R Studio and R programming language for Clustering**

Clustering data analysis tools for qualitative research available focused more on text-mining, which was not enough and appropriate to the current research work. R is a language for statistical computing and graphical needs (R Foundation, 2017). R provides a wide variety of statistical (classical statistical tests, linear and nonlinear modeling, time-series analysis, clustering) and graphical techniques, and is highly extensible (R Foundation, 2017).

R is free open source software, allowing anyone to use and modify as necessary. R is licensed under the GNU General Public License, with copyright held by The R Foundation for Statistical Computing. RStudio is the premier integrated development environment for R (R Project, the R Foundation, 2017). It is available in open source and commercial editions (R Project, the R Foundation, 2017). R is an integrated suite of software for data manipulation, calculation and graphical representation (R Project, the R Foundation, 2017).

Current research work implemented combined use of qualitative and quantitative methods, getting control over the data analysis process. While demographic representation of data used quantitative analysis, content analysis was fully deduced via clustering data analysis of qualitative method using “R Studio”. Researcher identified overarching categories and variables defined in constructs and built data with all related field variables in excel sheets for
each construct. These data related specifically for each construct was loaded into "R Studio" for cluster analysis.

Using this structure, Researcher proceeded to content data analysis for all the observations in each cluster identifying core common themes, using cluster membership. Researcher presented a discussion on list of observations on the selected clusters in the current research study and the analysis after returning to the original qualitative data observations and notes based on cluster memberships.

**Identification and Development of themes**

In-depth analysis of major themes occurred after the identification of themes in the clusters. Themes indicating a specific sequence of events were interconnected (Creswell, 2007). Themes and subthemes were identified following a grounded theory approach in the selected clusters for data analysis.

After coding and interconnection of themes, elucidation of the meaningful extractions of the experiences from the analysis of the data occurred through the application of reduction and imaginative variation techniques (Creswell, 2007).

Data classification into meaningful themes occurred to shed light on answers to the research questions and the successful start-up phenomenon as success factors. Using R Studio, experiences and perceptions are considered to develop wordclouds for each meaningful subtheme in major themes, so that a major pattern found with specific usage of words by the participants are highlighted.

The validation of findings of the current study occurred through the use of triangulation, and an external audit. A summary and narration of findings occurred following validation.